Investment Performance of a Concentrated Factor-Based Strategy

François Pierre Koninckx

Dissertation written under the supervision of professor Pedro Barroso

Dissertation submitted in partial fulfilment of requirements for the MSc in International Finance, at the Universidade Católica Portuguesa, 05/04/2023.

Abstract

I discuss the possibility to create an effective concentrated portfolio. The research starts with the observation that retail investors often fail to diversify their portfolios as much as advocated by traditional portfolio theory. Building on this observation, I create a factor-based concentrated trading strategy, to give a straightforward and relatively easy way to invest for retail investors. The strategy uses value, quality and beta as signals to construct monthly portfolios. Over the tested sample, the concentrated portfolio produces a significant alpha that is robust to the CAPM, and the Fama-French 3-Factor and 5-Factor models. Three portfolios are constructed, one long only, one long-short, and a naïve 1/N portfolio. The naïve portfolio is used as a benchmark of excess returns.

Keyword: Concentration, Factor investing, Portfolio construction, Under diversification.

Este trabalho discute a possibilidade de criar uma carteira concentrada eficaz. A pesquisa começa com a observação que investidores de retalho falham, muitas vezes, em diversificar as suas carteiras tanto quanto defende a teoria tradicional da carteira. Com base nesta observação, criei uma estratégia de investimento concentrado baseada em factores, para proporcionar uma forma simples e relativamente fácil para os investidores de retalho investirem. A estratégia usa valor, qualidade e beta como sinais para uma construção mensal de carteiras. A estratégia produz um alpha significativo ao longo da amostra testada que é robusto ao CAPM e aos modelos de 3 e 5 factores de Fama e French. Eu construo três carteiras: uma apenas longa, outra longa-curta e uma naive 1/N. A carteira naive é usada como benchmark.

Palavras-Chave: Concentração, Investimento em Factor, Construção de Portfólio, Diversificação Insuficiente

Investment Performance of a Concentrated Factor-Based Strategy: to what extent is it possible to produce an alpha with the creation of a factor-based trading strategy?

François Pierre Koninckx

Acknowledgements

I would like to thank everyone that directly or indirectly supported me through the creation and implementation of my project. I am grateful to Universidade Católica Portuguesa staff which provided me with the needed knowledge to start the project. I would especially like to thank my supervisor, Pedro Barroso who supported me with my ideas.

I would like to thank Roman Patris to have reviewed my work and made the needed change in syntax. I would like to thank Matilde Sepúlveda for the translation of the abstract. I am thanking my parents who have been of great support throughout my entire studies.

Table of Contents

Introduction

Nowadays, the first thing taught in portfolio theory is diversification. Markowitz (1952) show how to model and use diversification to improve portfolio performance in his influential paper, Portfolio Selection. In the findings, Markowitz (1952) shows that investors can drive away firm specific risk when investing in a portfolio composed of a diverse set of securities. Indeed, the mean-variance optimal portfolio offers the best risk adjusted returns an investor can aim for. Rational investors should follow the diversified composition when creating a portfolio. However, in practice, we observe that many investors diverge from the optimal allocation to their detriment.

Kelly (1995) investigates the extent to which investors are diversified in the US. The author finds that the median number of holdings in portfolios is one, demonstrating a clear lack of diversification at the investor level. Campbell (2006) and Ivković, Sialm & Weisbenner (2008) support the evidence and find that stock holdings for individuals range between two and 3.9 stocks. Similar behavior is witnessed at the fund level, even if the level of diversification is of greater extent.

Different reasons for the lack of diversification are often cited in the literature. In 2009, Boyle et al. show that information advantage at the investor level can be the driver of concentration. They show that superior information leads to the concentration of portfolios in a specific set of securities. Veldkamp & Van Nieuwerburgh, (2010) reach a similar conclusion while looking at investors' behaviors in the selection of securities. In addition, overconfidence of individuals is often cited as a driver of excessive trading behaviors (Barber & Odean 2013). Anderson (2008) find that overconfidence predicts the weight of risky assets in a portfolio. Furthermore, Chen et al. (2007) provide evidence that overconfidence leads to undiversified portfolios in China.

The evidence against diversification and its underlying reasons exists in practice. The proof of portfolio concentration leads us to think about ways to under diversify a portfolio in a rational and consistent manner. We have proof from investors like Warren Buffett that the market can be outperformed consistently with under diversified portfolios. Documentation exists about their investment beliefs. Buffett's Alpha paper by Frazzini, Kabiller, and Pedersen (2013) investigate Warren Buffett investment strategy. In the research, the authors seek to explain the drivers of Berkshire Hathaway's alpha with explanatory variables. Frazzini, Kabiller, and Pedersen (2013) find that the investor alpha can be pinned down to a few factors. The combination of leverage and the investment in cheap, safe and quality stock seems to explain the returns.

This findings in Frazzini, Kabiller, and Pedersen (2013) research led me to the following problem statement and sub-questions:

Investment Performance of a Concentrated Factor-Based Strategy: to what extent is it possible to produce an alpha with the creation of a factor-based trading strategy?

- Is it possible to replicate Warren Buffett's strategy in a simple way?
- Can a factor-based concentrated strategy beat a naïve portfolio?
- Can I extract significant returns with an undiversified concentrated portfolio?
- Can I extract an alpha with a factor-based concentration strategy?

In order to find an answer to my problem statement and related sub-questions, the thesis will be structured as follows.

First, I start with a review of the literature. I introduce modern portfolio theory and the models of Markwoitz (1952) and Sharpe (1964). I look at the rational behavior of investors in practice and the rules that should be followed to obtain a mean-variance efficient portfolio. I continue the section by looking at portfolio concentration. I gather proof at the investor and fund levels that concentration exists in practice. I then dig into the reasons for diversification. I find that superior information and overconfidence are the main drivers of under diversification. Second, I dig into the research design. In this section, I aim to introduce my model. I start by looking at the model of Frazzini, Kabiller, and Pedersen (2013), from where my idea emerged. That is, from the paper, I obtain the three factors that I include in my strategy. I continue by elaborating on the three factors I chose, what metrics define them and the reason why I decided to pick them. In this section, I build the ground for my strategy by explaining the logic behind it. Third, I look at the data. The data section introduces the US dataset, which ranges from 1970 to 2022 and is composed of NASDAQ, NYSE and AMEX securities. In this section, I describe the cleaning process I go through and how I deal with the different adjustments I make in the dataset. Fourth, I look at the methodology. In this section, I introduce the main strategy, the long only factor-based strategy. Here I explain the construction of the different portfolios and how I use the factors in that regard. I also explain the alternative strategies I create. Fifth, I introduce the expected hypothesis of my work. This means that, in this section, I create five hypotheses that I test in the results section. Sixth I present the results of my analysis. Here I lay down my findings for the different strategies. I describe the key results of the factor-based concentrated strategy, the long-short, the naïve portfolio and market returns. I finish the section by looking at the results of the different statistical tests. I find that the factor-based concentrated strategy can extract significant returns and produce an alpha. I enclose the research by elaborating on the limitations of these results and the place there is for further research.

In my research, I must note that I do not aim to provide proof against portfolio diversification. With the strategy created, I do not aim to show that concentration is more efficient than diversification. I simply aim to present a straightforward and reliable portfolio construction method, focusing on concentration. From the observation that investors are poorly diversified agents, I aim to give a simple way of investing in a small number of securities. To my knowledge, I am the first to implement the given strategy with the given method. Bias and caveat exist in the project. I aim to pave the way for reliable methods to concentrate a portfolio. In addition, I create a "concentration" strategy. The strategy averages 22 stocks per portfolios. This composition can be described as diversified. However, I only select 1% of securities, which I consider as an undiversified allocation in the research.

In my text and research, all the results are in excess returns. The term return is used interchangeably with excess returns for the sake of redundance. I describe the concentration factor-based strategy as under diversified portfolio, long only, concentration strategy. All the terms point to the same strategy.
Knowing what the thesis focuses on I can start off with the thought process which led me to create my strategy.

1. Literature Review

In this section, I aim to cover the theory that leads me to understand the rationale behind the trading strategy I create. To fully understand the choices, I decide to split the review of literature in two parts. The first part encompasses portfolio theory. I start by covering traditional portfolio theory and diversification, pioneered by Markowitz's work in 1952. I continue the section by understanding the theory of Fisher (1964) and the Capital Asset Pricing Model (CAPM). The second section focuses on portfolio concentration. I start to look at different studies and see that investors deviate from the traditional mean-variance optimal portfolio. I gather proof at the investor and fund levels. I then investigate the reasons why investors deviate from an optimal portfolio and cover two main reasons for the deviation. I see that some investors depart from the advocated choice, because they believe or have superior information, or are subject to behavioral bias such as overconfidence.

1.1. Traditional Portfolio Theory

The concept of diversification is one of the first concepts investors learn when discovering how to create a portfolio. "Do not hold all your eggs in the same basket" can sum up the logic very well. In 1738, Bernoulli was the first to introduce diversification in an economic context; however, without giving any mathematical foundation to it. He wrote, "*Another rule which may prove useful can be derived from our theory. This is the rule that it is advisable to divide goods which are exposed to some danger into several portions rather than to risk them all together*" (Bernoulli, 1954). However, no quantitative ground was set until 1952 when Markowitz first introduced, with his influential paper "*Portfolio selection*", the modern portfolio theory. From that point on, a large range of research and experiments have been conducted to showcase the benefits of a diversified portfolio.

I now set the grounds of modern portfolio theory and look at two of the most influential works.

1.2. Modern portfolio theory

With the introduction of the "Mean-Variance" model, Markowitz (1952) presents a method of diversification for investors. In the paper, Markowitz (1952) creates a logic to determine an optimal portfolio composition. This logic considers the expected returns of individual assets and their risks in the portfolio. The optimal composition of a portfolio is one that maximizes the expected returns for a given level of risk.

To that end, a tradeoff between expected return and risk is introduced. That is, an investor can attain higher expected returns; however, only by increasing the risk of the overall portfolio. This risk-return tradeoff and optimal combination is found by using a mathematical model made of three components: the expected return, the risk, gauged by the standard deviation, and the covariance of individual securities (Markowitz, 1952). The model suggests that, "when securities are combined in a portfolio, the risk involved in holding them together is lower than the sum of their individual risks" (Markowitz, 1952). This yields the grounds for diversification and diversified portfolios.

An efficient, diversified portfolio is one that either maximizes the expected return given the level of risk or one that minimizes the level of risk, given the expected return. In *Figure 1* below is depicted the feasible set of portfolios. The x-axis represents the risk of the portfolio, while the y-axis shows the expected returns. The feasible set of portfolios consists of any combination of securities that yields a portfolio weight of 1. In addition, the figure depicts the efficient set of portfolios, represented by the blue line. This line illustrates the solution to the optimization problem presented by Markowitz (1952). The efficient set of portfolios is called the efficient frontier and represents the change in risk-reward ratio that a diversified investor can obtain.

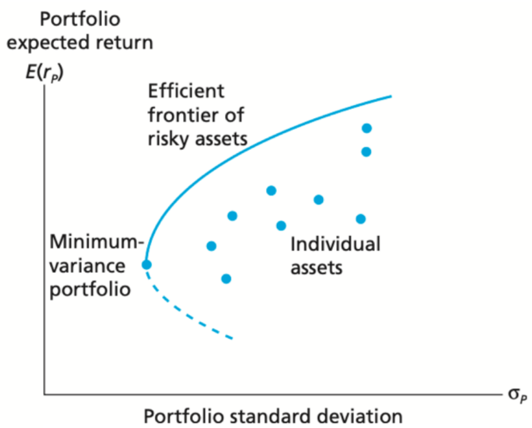

Figure 1: Investment opportunity set

Efficient frontier

1.3. CAPM

In 1964, Sharpe introduces a new framework, which builds over Markowitz's idea. Sharpe's (1964) framework introduces three important assumptions for the model to hold. He assumes the rationality of investors, homogenous expectations about market information and that investors can borrow and lend at risk-free rate. In addition, investors have the ability to short sell without limits (Sharpe, 1964)[1].

In his influential work, Sharpe (1964), introduces the Capital Asset Pricing Model. The model paves the way for investors to determine the expected return of individual assets or portfolios, based on the risk and correlations the assets have with the market. In addition, the introduction of the risk-free rate introduces the Capital Market Line (CML). The CML is a tangent to the efficient frontier. It introduces an efficient combination between investment in the risk-free rate securities and the market portfolio. Mean-variance optimizers will only hold portfolios along the CML line (*Figure 2*). The CML is depicted as the Capital Asset Line (CAL) in *Figure 2*, terms can be used interchangeably.

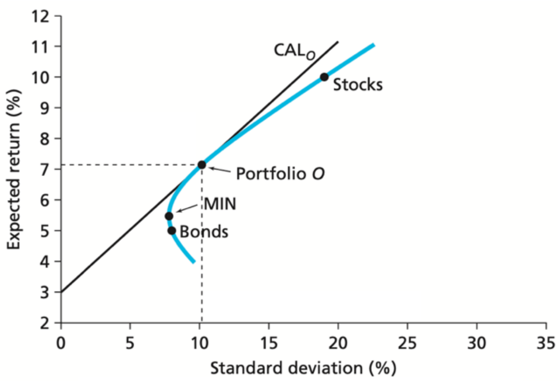

Figure 2: Capital Market Line

The equation of the CML and the CAPM model start with the assumption that investors are averse to risk. That is, investors require additional return for investing in assets riskier than the risk-free rate. This additional return requested by investors is called the risk premium. The risk

[1] Sharpe (1964) is cited, in this section as the main contributor of the CAPM. However, the model is a work of joined forces. Sharpe (1964), Treynor (1962), Lintner (1965a, b) and Mossin (1966) are the pioneer of the CAPM.

premium is the measure of an asset's sensitivity to market returns. That is, the risk premium is measured by the beta of securities. The beta is calculated as the covariance of the given asset/portfolio with the market, divided by the variance of market returns (Bodie, Kane & Marcus, 2018). Given the beta obtained, it is possible to derive the expected return of an efficient portfolio with the following equation:

$$\beta_i = \frac{Cov\ (r_i\ , r_m)}{Var\ (r_m)} \tag{1}$$

$$E(r_i) = r_f + \beta_i\ (E(r_m) - r_f) \tag{2}$$

Where $E(r_i)$ is the expected return of security i, r_f is the risk-free rate, β_i is the risk premium of individual securities and $E(r_m)$is the expected return of the market.

The CAPM model allows investors to include the dependance of assets with the market. This yields a model which gives an efficient allocation of assets given a risk-return optimization setting. In addition, "CAPM proves that the risk of individual assets or portfolios is not only related to their own characteristics but to the overall market as well" (Bodie, Kane & Marcus, 2018). It shows that idiosyncratic risk can be driven away; however, on the other hand, systematic risk cannot (Sharpe, 1964).

Modern portfolio theory, introduced by Markowitz (1952) and Sharpe (1964), pave the way to construct an optimal portfolio. That is, with their theory, the two authors show how a mean-variance rational investor should behave. Indeed, in theory, no deviation to the model should exist. The reason is that extra risk is taken for a given level of returns if deviation from the pattern exists. However, many investors do not follow the strict principle of diversification and mean-variance optimization. For the majority, it results in poor investment decisions and returns over time. Nevertheless, a small portion of individuals deviate from diversification and still extract abnormal returns. They extract their returns from a narrow selection of securities to build their portfolio. The riskier portfolios, if constructed correctly, can provide abnormal returns. Investors like Warren Buffett, Charlie Munger, Peter lynch or Terry Smiths are a few well known investors who extract additional returns by portfolio concentration.

I will now introduce papers that tackle portfolio concentration. I will start by highlighting the fact that investors are poorly diversified agents and will look at mutual funds as well. I will then see that information advantage drives some investors to concentrate. I will continue to stress a behavioral bias – overconfidence – which can lead to concentration.

1.4. Portfolio Concentration

I now take a contrasted view from traditional portfolio theory and look at portfolio concentration. Before I dig in, I have to note: I know that diversification yields the best risk-adjusted returns. I do not aim to provide counter proof. Many academics have investigated the matter and the consensus points towards the net benefits of diversifications[2]. Nevertheless, this section intends to showcase two points. First, I show that a lack of diversification exists at the retail and institutional level, in practice. Second, I summarize the reasons why investors are not mean-variance efficient. Understanding the logic behind portfolio concentration sheds light on investors' decision-making process.

After I understand the rationale behind concentration and the reasons around it, I introduce a trading strategy centered around concentration. My strategy intends to propose a straightforward method of investing for individual investors. The strategy aims to obliterate bias usually showcased by investors. I then test the model and draw conclusions on its efficiency.

1.4.1. Under diversification

"If finance textbooks are to be believed, investors hold an equally weighted portfolio of twenty or more stocks. By doing so they eliminate the idiosyncratic risk and face only undiversifiable risk, equal to the average covariance of returns of the portfolio" (Kelly, 1995). However, in practice, individuals depart from the traditional composition. Some investors believe they have an information advantage. Others face bias which leads them to depart from the optimal allocation (Liu, 2008).

[2] See, for example, Marston (2011), Statman & Scheid (2008), Levy & Sarnat (1970), Driessen & Laeven (2007), Evans & Archer (1968) or Lhabitant (2017), for studies covering diversification.

1.4.2.2. Concentration at retail level

This section aims to illustrate the gap between strict theory and practice, at the investor and fund levels. I first point out that investors are under diversified agents. Where theory predicts mean-variance optimal portfolios, practice shows a deviation. Then, I dig into the reasons that explain the lack of diversification.

Kelly (1995) investigates how diversified investors are in the US, using a sample of 3665 households to conduct two analyses. The first survey was conducted in the 1960s with a median of holding of $3,800. The second, organised in 1983, considered high income participants (more than 100,000$ of yearly income). For the 1960s sample, Kelly (1995) found that the median number of stocks held in a portfolio was one. Comparably, in the high-income portion, the median of stocks in individual portfolios was 10. These results demonstrate that US investors hold a less than optimal portfolio. This being said, the surveys were performed more than half a century ago. Since then, ETFs were introduced and widely adopted and barriers to investing have decreased[3]. However, even if the results need to be taken lightly, it shows that investors do not always hold optimal portfolios. The case applies for wealthy investors as well. In 2006, Campbell confirms Kelly (1995) evidence by finding that until 2001, the median number of stock holdings for individual investors was two and rose to three thereafter. He noted that the findings did not consider mutual funds stock holdings, which would increase the median holding. However, when it comes to individual holdings, it shows that investors are not mean-variance efficient in their decisions. In 2004, Goetzmann & Kumar confirmed that retail investors are under diversified by conducting a study with 60,000 individuals. The sample ranged from 1991 to 1996 in the US. They found that many investors of their sample were under diversified. Ivković, Sialm & Weisbenner (2008) reached a similar concentration conclusion by finding that typical individuals hold a portfolio of 3.9 stocks.

1.4.2.2. Concentration at fund level

Despite the broad proof that passive investment yields significantly better returns than active investment (Carhart (1997), Daniel et al. (1997), Gruber (1996), Malkiel (1995), Jensen (1968)), often characterized by under diversification, I focus on concentration in the mutual fund industry for the sake of the strategy I create. In this section, I do not try to bring proof

[3] I.e. Cost of investing have decreased. For example, brokers as Robinhood or Interactive Brokers offer free trading on US securities. Cost of information have decreased with the spread of online news.

against passive management strategies. However, I seek to demonstrate that a share of institutional investors going for portfolio concentration exists. The focus of this section is not on the results obtained from concentration. In other words, the reader should take the results, regardless of their direction, not as their superior performances, but as proof that concentration exists at the fund level.

In an attempt to understand the driver of portfolio concentration, Kacperczyk, Sialm, & Zheng (2005) uses mutual fund data from 1984 to 1999. They construct portfolios of funds classified by their levels of concentration in industry. They find that mutual funds with a concentration above the median yield a return 1,22% higher than funds with a concentration below average (1,58% vs 0,36%). The results hold after expenses. In another study, Fulkerson & Riley (2019) investigate fund performance and concentration. They conduct a US based study. They start from the view that every fund should be fully diversified and invest in a portfolio with no idiosyncratic risk. However, as managers acquire information about certain securities, the portfolio composition should shift. Indeed, if managers have superior information, the concentration in the given securities should increase. They find that when a fund increases its concentration by one standard-deviation, risk adjusted returns increase by 0.24%, on average. Fulkerson & Riley's study (2019) is similar to the one by Treynor and Black (1973) for individuals. The above-mentioned studies quantify concentration using a Herfindahl–Hirschman Index-like method. Choi et al. (2017), gather data from security holdings of 10,771 institutional investors portfolio, to look at the relation between returns and concentration. They find that concentration in home country or foreign country industry is associated with higher risk adjusted returns. This empirical finding goes hand in hand with the model by Veldkamp & Van Nieuwerburgh, (2010). Sapp & Yan (2008) conduct a study with US fund data from 1984 to 2002. They look at gross return of funds given the number of securities held in their portfolio. In their research, they do not find evidence that the concentrated funds outperform the diversified ones.

With the different studies covered, I have evidence that concentration exists at the fund level. I now dig into the reasons why investors deviate from an optimal portfolio. I start with information advantage. I continue the section with behavioral bias.

1.4.2. Reasons behind concentration

I have now viewed that investors lack diversification in their portfolios. In other terms, in practice, investors are not rational. In this section, I look into the reasons behind the seemingly irrational choices of investors. I first analyze information advantage, then behavioral bias.

1.4.2.2. Information advantage

In 1973, Treynor and Black take a new view on the optimal portfolio. They state that an investor with no information advantage should hold a diversified portfolio, as advocated by Markowitz (1952). However, in case the investor has valuable information about specific securities, he should act upon it. Thus, his information advantage should be reflected in the weight of the given securities, in the portfolio.

Boyle et al. (2009) take Treynor and Black's (1973) reasoning one step further and create a model of ambiguity for investors. The starting point of their research is Markowitz's model (1952). From there, they note that, in practice, it is challenging to forecast expected returns, as advocated by Merton (1980). To challenge the model, an ambiguity factor is introduced. Ambiguity equals the information investors have, i.e., historical prices of securities. This ambiguity can be increased or decreased with an information advantage. Investors are averse to ambiguity. Boyle et al. (2009) look at the behavior of investors for portfolio creation when the level of ambiguity is known. They find that, the higher the familiarity towards an asset, the more the portfolio will be skewed towards it. That is, if investors have an information advantage over a set of securities, their portfolio composition will be skewed towards them. Uppal & Wang (2003) confirm that change in ambiguity leads to portfolio concentration.

Veldkamp & Van Nieuwerburgh, (2010) try to rationalize the choice of concentration. In their approach, they solve a joint problem of information specialization and investment. Their starting point is formulated as follows: "Specialization arises because the more an investor holds of an asset, the more valuable it is to learn about that asset; but the more an investor learns about the asset, the more valuable that asset is to hold" (Veldkamp & Van Nieuwerburgh, 2010). They find that, as in Boyle et al. (2009), investors concentrate their holdings in assets they learn about, more than predicted in standard theory. Their choice is driven by a willingness to resolve uncertainty. The view is interesting as uncertainty mimics risk in a certain way. "The main message is that when investors can choose what information to acquire before they invest, they may invest in portfolios that would be sub-optimal for an investor who has not learned. From the point of view of standard portfolio theory, these portfolios might be deemed anomalous or irrational" (Veldkamp & Van Nieuwerburgh, 2010). It shows that irrationality might not be the driver of portfolio concentration as advocated by traditional portfolio theory. In an attempt to look at the home bias in portfolios, Hatchondo (2008) conducts a study looking at investors' decision making. In the study, investors are faced with investment decisions between local and foreign firms. Each investor has a different information set. They observe that investors opt for local investment. In addition, investors choose to concentrate their portfolios in firms of their regions. In another study, Huberman (2001) looks at a specific company's employees and their investment behavior. They find evidence that investors are more prone to invest in local businesses, which pushes towards the concentration of investment.

Ivković, Sialm & Weisbenner (2008) investigate the reasons behind concentration. They state that if undiversified portfolios exhibit a superior performance compared to diversified ones, it must be linked to stock picking abilities and information advantage. The authors use a sample of 78,000 household trades and monthly positions from 1991 to 1996, across various income classes. Three results are obtained. First, the investments made by concentrated investors significantly outperform diversified ones, by less than a percent, over the year following the purchase, overall. Second, when portfolios are larger than 25,000$, the difference increases to three percent. Third, the difference in returns vanishes with small portfolios, i.e., less than 25,000$.

Their results point towards the ability of investors to extract an information advantage. Nevertheless, the conclusion should be taken lightly. The authors qualify a portfolio as diversified if it has three or more securities. Even if the level of diversification depends on which assets are included in a portfolio, three securities have a high chance of producing a concentrated portfolio. Given the mixed results of their analysis, these should be taken lightly. Nonetheless, they indicate that information advantage could be a driver of concentration.

I have tackled a possible reason for the concentration of portfolios. I shall now look at investors' behavioural bias as a potential reason for under diversification.

1.4.2.2. Overconfidence

In the previous section, I first saw that some investors are undiversified agents. This under diversification is in part driven by the belief in an information advantage. These pieces of information are then incorporated within investor preferences while constructing a portfolio. In addition to that, a behavioral bias exists, which prevents investors from fully diversifying: overconfidence. I see, in this section, that most of the papers covering the subject link overconfidence to excessive trading from individuals. However, to my knowledge, fewer studies have investigated the direct link between overconfidence and under diversifications. Nonetheless, overconfidence leads investors to exhibit irrational behavior. Irrational behavior then leads to overweight securities for which investors are confident about. I examine the major research regarding overconfidence at the investor level. I then make the indirect link between overconfidence and under diversification.

A large body of literature in psychology led by Moore and Healy (2008), Odean (1999), Alpert and Raiffa (1982), Svenson (1981), Fischhoff, Slovic, and Lichtenstein (1977), find that individuals are overconfident in their abilities. This overconfidence, driven by the belief in being better than average, results in behavioral bias when it comes to investing. It is found that individuals are driven by the idea that their beliefs will drive them to better investment decisions when creating a portfolio. This results in high trading turnover and suboptimal portfolio construction (Hvidkjaer, 2008). Following this view, Barber & Odean (2013) states that individual investors behave differently than advised in traditional finance models and relate it partly to overconfidence. They say that “Most individual investors hold under diversified portfolios. In addition, apparently uninformed investors trade actively, speculatively, and to their detriment” Barber & Odean (2013).

In their first attempt to tackle the subject of overconfidence at the investor level, Barber & Odean (2000), define overconfidence as “the overestimation of the value of private information”. This overestimation causes higher than optimal trading and thus leads to sub optimal returns. They find that investors in the US trade too much. In another research, Barber & Odean (2001) look at the matter, but this time by observing gender behavior. They build their research on the assumption that men are more confident than women. They find that men trade 45% more than women. This leads to an average decline in returns of 1.72% for women and 2.65% for men. Their findings confirm their hypothesis, according to which overconfidence exists in financial markets. Barber & Odean (2001) confirm the belief of their previous study as they find that "an overconfident investor overestimates the precision of his information and thereby the expected gains of trading" Barber & Odean (2001). In a similar fashion, Scheinkman & Xiong (2003) look at overconfidence and its relation to market bubbles. According to Barber & Odean (2000, 2001), overconfidence is modeled as the belief in superior personal information compared to other market participants. In this model, they find that participants will buy securities at a price higher than their value. This behavior is led by overconfidence. Agents believe they will, in the future, be able to sell at a higher price. This leads them to overbid. Barberis and Thaler (2002) find similar results about overbidding for securities. Penga & Xiong (2006) point out that overconfidence is particularly severe in tasks where a significant share of judgment is needed. The case is strong for investors when it comes to portfolio construction.

In the previous research gathered, the link is made between confidence and trading behaviors. However, the literature lacks links between overconfidence and portfolio construction. Anderson (2008) attempts to make the tie between overconfidence, trading and diversification. In Anderson (2008) model, investors are optimistic about their stock picking capabilities. These abilities lead investors to deviate from an optimal portfolio. Anderson (2008) finds that overconfidence predicts the share and weight of risky assets included in a portfolio. Grinblatt and Keloharju (2009) also look at the link between confidence and concentration. They find that overconfident investors are prone to form portfolios with excessive weights. These excessive weights are placed on private information gathered by investors. Chen et al. (2007) conduct an analysis on Chinese investors' behaviors. They find that the Chinese feel certain about their investment skills and knowledge. This overconfidence leads them to trade too much and hold undiversified portfolios.

I now have an overall view on the bias that leads investors to invest in a concentrated portfolio.

2. Research Design

This section aims to introduce the factor-based concentration strategy. I first start to introduce where I got the idea of a concentrated strategy. I then explain the reasoning behind my choice and how I decide to implement it.

2.1. A factor-based concentration strategy

The idea of a factor-based concentration strategy comes from empirical and practical observations. An existing body of literature provides extensive proof that passive investing outperforms active investing on average, in the long run (Carhart (1997), Daniel et al. (1997), Gruber (1996), Malkiel (1995), Jensen (1968)). Proven to be true and tested multiple times, some investors are however able to perform consistently without investing in the market portfolio. That is, names like Warren Buffett, Charlie Munger, Peter Lynch or Terry Smith, are often heard and cited as the greatest investors of all times. These investors have indeed been able to implement strategies, yielding significant returns over time.

The common feature between these fund managers is the focus on a restricted number of securities and the creation of concentrated portfolios. Their portfolios have a common feature. They combine securities with strong fundamentals and growth potential. In addition, strong due diligence is performed before the investment period, to gauge the stability of a security. Through this process, these investors have been able to construct portfolios with a limited number of holdings that perform over time.

From the given observation came the idea to find a way to replicate their strategies. Having limited resources and analytical power, I aim to create a simplistic model. This model aims to group, in a simple and efficient way, the investment beliefs of these successful fund managers into factors. The factors attempt to offer a quantitative way to mimic these investment beliefs.

The model I are proposing is, to my knowledge, the first of the sort. It entails that the expected returns from the strategy are unknown and can go in either direction. I expect to create a simple model, replicable by retail investors, which yields positive returns over time. However, in practice, the strategy might yield positive or negative returns.

In the next section, I introduce the model and the three factors I picked to replicate Warren Buffett's investment beliefs. I explain the rationale behind it.

2.2. Warren Buffett replication – A three factor composition

The concentration portfolio model is built on the investment beliefs of Warren Buffett as well as his investment style. The starting point of the model is the analysis realized by Frazzini, Kabiller, and Pedersen in 2013. In their paper "Buffett's Alpha", they investigate the alpha of Berkshire Hathaway. Throughout their process, they acknowledge that Buffett has achieved significant returns over time with a Sharpe ratio of 0.76. In addition, they highlight evidence that Warren Buffett Berkshire Hathaway has consistently extracted an alpha through time. They seek to understand what made that abnormal return realizable.

Throughout the analysis, they find that the generated returns are partly related to the ability to leverage the investments and invest in cheap, safe, quality securities. To test this, Frazzini, Kabiller, and Pedersen (2013) create a regression and seek to explain the origin of Buffett's alpha. They find that the abnormal returns become insignificant when alpha is controlled for beta, value and the quality factor. Indeed, the three factors can partly explain the alpha extracted by Warren Buffett. This is the starting point of my analysis.

I know, from Buffett's interviews and letter to shareholders (Buffett (2022), Winters (2019)), the investment beliefs that drive Buffett's investment decision. However, "Buffett's Alpha" (2013) provides empirical proof of the alpha's drivers. Based on their findings, I decide to pick three factors, which would replicate Warren Buffett's investment style. To that end, I aim to provide a relatively easy strategy to implement, which has proven to be efficient through time, for retail investors.

The regression used by Frazzini, Kabiller, and Pedersen (2013) to explain Buffett's alpha tests the monthly excess returns of Berkshire Hathaway against the market returns, size, value, momentum, beta and quality factors. Not all factors are significant in the model (appendix, *Figure 1)*. From that observation, I decided to only take the significant factors from the regression and create a strategy based on them.

From the output, the market, value, beta and quality factors are significant. I have interest in the factors which are company specific. That is, I do not to include the market factor in the creation of my strategy. I keep the value, beta and quality factors.

In the next section, I briefly introduce the three factors, their efficiency in predicting returns as well as the influential work done around them. I then explain how I combine the factors to create a concentrated factor-based strategy.

2.3. Reasoning behind the three factors

2.3.1. Beta factor

In Buffett's alphas (2013), Frazzini, Kabiller, and Pedersen decide to control for a beta factor. In the analysis, the beta factor is the "betting against beta factor" (BaB) of Frazzini and Pedersen (2014). The authors look at the alpha that produces low and high beta stocks. They find that "investors bid up high beta assets. This makes high beta assets low alpha securities. In addition, they find that going long in low beta assets and short in high beta assets, produces significant positive, risk adjusted returns" (Frazzini and Pederson, 2014). It is known that low beta assets are associated with "safer" securities, i.e., their movement with the market is of lower magnitude than high beta assets. Furthermore, Warren Buffett's investment style seeks to invest in safe assets. To that end, Frazzini, Kabiller, and Pedersen (2013) decide to control for the BaB factor and see if it can explain part of Warren Buffett's alpha.

While running the test, Frazzini, Kabiller, and Pedersen (2013) find that the BaB factor can significantly explain Buffett's returns. That is, the factor points towards an investment strategy which overweighs low beta stocks to produce returns. In Buffett's Alpha (2013) paper, the factor is significant at any conventional level (t=2.67). I select the factor to create the portfolios.

2.3.2. Value Factor

Graduating from Columbia Business school in 1951 (Columbia University, 2014), Warren Buffett was student of Benjamin Graham, the pioneer of value investing. Having started a business venture together (Frazzini, Kabiller, and Pedersen, 2013), it is widely known that Graham was of considerable influence to Buffett's investment style. In effect, Warren Buffet drives his investment decisions by selecting value stocks, as advocated by Graham and Dodd's (1934) book. In the book, the authors advocate investing in "cheap" companies with strong fundamentals. "They argue that investment decisions should be influenced by the security intrinsic worth, rather than the fluctuation in price incurred by market movements" (Graham and Dodd, 1934).

Knowing that Warren Buffett is largely influenced by value investing, it is only logical for Frazzini, Kabiller, and Pedersen (2013) to use the factor to explain the excess returns of the fund manager. Indeed, in their analysis, value is controlled for. It is proven that a positive relation exists between investing in value and stock returns, as documented by Rosenberg, Reid, and Lanstein (1985), along with Fama and French (1993). By testing for value, Frazzini, Kabiller, and Pedersen (2013) find that the returns of Buffett are associated to those of the value factor. For that reason, I include a value factor in my model.

2.3.3. Quality Factor

In their quest to explain Buffett's alpha, Frazzini, Kabiller, and Pedersen (2013) introduce the quality factor. The quality factor represents profitable, stable, growing and high payout stocks. This factor is in line with Buffett's investment belief, as it targets superior and stable stocks having potential to yield significant returns in the future (Buffett 2022). Controlling for quality in their analysis is logical. A positive quality factor shows that quality stocks outperform junk stocks in general, as advocated and proven by Asness, Frazzini, and Pedersen (2013). Indeed, in their analysis, they find significant evidence that quality stocks outperform junk stocks.

Included in the regression which aims to explain Buffett's alpha, the quality factor is significant. This shows that Buffett's portfolios are constructed by considering the quality of assets. Therefore, I decide to include this third factor my model.

I covered the different factors which explain Buffett's alpha. Based on these three significant factors, I have decided to build an investment strategy. Before I dig into the methodology and explain how the investment strategy was created, I elaborate on my data choice and how I controlled for the data cleaning process.

3. Data

The strategy is applied on US securities only. This choice is supported by three reasons. First, data on US individual securities are accessible to student student from data vendors. Second, Kenneth French's data library is used and only covers US based securities. Third, the strategy is implemented from the point of view of an individual investor. I aim to implement a low transaction costs strategy. The only securities benefitting from commission-free trades yet are US securities[4]. To perform my analysis, data from all NYSE, NASDAQ and AMEX listed securities from January 1970 until December 2022 are retrieve from CRSP and Compustat. In addition, Kenneth R. French's database is used. CRSP's data base is used for monthly securities prices. Compustat retrieve accounting data. Kenneth R. French's data library is used to have the market returns as well as the risk-free rate. The sample period choice is made based on the availability of data on WRDS. No accounting data are available prior to 1970.

Data retrieved on individual securities are monthly stock price (PRC). In addition to the monthly price, I gather the adjustment factor (CFACPR), used to convert monthly price into adjusted monthly price[5]. Ticker symbols (TICKER) are collected as well as the exchange code (EXCHCD) and the permno code (PERMNO)[6]. For accounting data, I used the Compustat database and included the following ticker, permno code, book-to-market values (bm) as well as a gross profit measure (GProf)[7]. The third dataset used is the Kenneth R. French data library. I collected the excess market returns (Mkt-RF), small minus big factor (SMB), high minus low factor (HML) and the risk-free rate (RF).

3.1. Data Cleaning Process

Data cleaning and the adjustment process are a significant part of the work. In total, more than almost 7 million rows of data[8] are processed, making the cleaning process significant and challenging. Three different adjustments are made to get data which is ready to work with. These are individually described.

[4] For example, Robinhood or Interactive Brokers offer commission-free trade over US securities only.

[5] The adjustment factor account for dividends, split and right offering.

[6] Permno code is a unique security code which allows to merge datasets between CRSP and Compustat.

[7] Defined as gross profit divided by total assets

[8] 2,919,530 rows for the accounting dataset and 3,994,626 rows for the price data

3.1.1. NA Values

The dataset suffers from missing values in the price and adjustment factor columns. Overall, 43,269 rows had missing prices and adjustment factor values. Missing prices and adjustment factor are always paired. To deal with the matter, I replace a missing value with the previous one. The transformation decreases the accuracy of price change, as 43,269 individual price changes are eliminated. However, the transformation should not impact the results significantly as 1.08% of price points are accounted as missing values. In addition, no group of missing values over multiple months for a given stock exists. The accounting dataset accounts for missing values at the factor level as well. The value factor has 143,228 rows of data missing, accounting for 4.9% of missing values. The quality factor has 75,131 missing values (2.57% of missing values). I deal with missing observations by replacing NA with the previous rate. I have to note that the given adjustments will reduce the accuracy of my results.

3.1.2. Adjustment factor

In the dataset, the adjustment factor is provided to incorporate stock splits, dividends and rights offerings into a stock price. In the CRSP dataset, the "Data Description Guide" states that I should divide individual security prices with the adjustment factor to control for corporate actions (CRSP, 2020). The adjustment factor ranges from 0 to 1215. When I automate the process of calculation, an error occurs in my code, as dividing by zero yields undefined results. In order to control the matter, I changed stocks having zero values in their adjustment factor column by one. In my dataset, 45,001 values display the problem. The operation is done to avoid undefined security prices.

3.1.3. Outliers

In the data set, outliers exist. To obtain a decent analysis, I have to mitigate their effects. In order to do that, I decided to erase stocks with values of less than 5$, as done in Jegadeesh & Titman (1993), for example. Different techniques exist to control for outliers; I decided to implement said technique.

3.1.4. Merging data sets

In my research, three datasets are used. Each dataset has its specificities and have to be merged. The price and accounting datasets are joined with the permno code. That is, each security is assigned a permno code. This code is unique to each security regardless of its listing or delisting. What's more, some ticker symbols are reused for new listings when a security is delisted. With the permno, I avoid pairing two different securities into one (WRDS, 2020). In addition to merging securities with the permno code, dates are used as well to match the price and accounting data. The market data (Kenneth R. French dataset) is matched along with the dates to the CRSP and Compustat datasets.

4. Methodology

In this section, I lay out the methodology of my strategy. I start off with the description of the returns. I then detail the use of the factors to construct portfolios.

To obtain the average return of my strategy, I opt for the following method to calculate returns. The first step in calculating the returns[9] is to convert monthly prices into monthly returns, for the individual securities. The monthly arithmetic returns are defined as:

$$r_{i,t} = \frac{P_{i,t} - P_{i,t0}}{P_{i,t0}} \tag{3}$$

Where $r_{i,t}$ is the monthly arithmetic return for a given stock *i* at time *t* and where $P_{i,t}$ is equal to the price of stock *i* at time *t*. The risk-free rate is then subtracted to obtain monthly excess returns.

I then calculate the beta of individual securities. The beta of individual securities is computed following the methodology of Bodie, Kane & Marcus (2019) and the sample size of Liu, Stambaugh, & Yuan (2018). Liu, Stambaugh, & Yuan (2018) state that "there are numerous approaches for estimating betas on individual stocks, and the literature does not really offer a consensus". To that end, I decide to take the standard beta approach from Bodie, Kane & Marcus (2019) and apply the look back period used in Liu, Stambaugh, & Yuan (2018) and Frazzini & Pedersen (2014). This means I estimate the beta by regressing the individual excess returns of securities over the excess market returns. I picked a window size of 60 months. In

[9] Returns are adjusted for split, dividends and right offering.

the analysis, the beta of a security is not calculated if it does not have 60 consecutive observations of returns. If a security does not have a beta, I do not consider it for the portfolio creation. I proceed as follows:

$$r_{E(i,t)} = \alpha + \beta_i \, r_{E(m,t)} + \varepsilon_{\,i,t} \tag{4}$$

Where $r_{E(i,t)}$ is the monthly excess return of security *i* at time *t* and where $r_{E(m,t)}$ is the monthly market excess return *m*, at time *t*.

I continue by computing the value factor. As a proxy for value, I take the book-to-market ratio of firms as in Fama & French (1993). The book-to-market factor of individual firms is as follow:

$$\frac{Book\ value\ of\ equity_{i,t}}{Market\ value\ of\ equity_{i,t}} \tag{5}$$

Where I retrieve the ratio for each security *i,* at time *t*, at a monthly level.

The last factor selected is the quality factor. I proxy the quality of a firm with the following measure:

$$\frac{Gross\ profit_{i,t}}{Total\ assets_{i,t}} \tag{6}$$

The quality factor is extracted for each security i, at time t, at a monthly level. However, for public companies, profits and total assets are often only available at quarterly level. If that is the case, the previous quarter value is selected for the monthly increment.

In Buffett's Alpha, the methodology of Asness, Frazzini and Pedersen (2013) is used. Being limited by the accounting measures available in WRDS, I am not able to replicate the methodology. I then opt for the measure used by Novy-Marx (2013) to mimic the quality of a firm. In their research, quality is proxied by the ratio of gross profitability to the total assets of a company.

Given each firm's monthly returns and their respective accounting factors, I can proceed to the ranking. Each firm is ranked three different times. First, based on individual betas. The firm with the lowest beta is ranked number one and the one with the highest beta is ranked last. Second, firms are ranked based on value. The securities having the largest book-to-market value are ranked first. Third is the quality factor. Firms with the highest ratio obtain the top ranking.

The securities individual ranking is then aggregated. To aggregate the three rankings into a single measure, the factor loadings from Frazzini, Kabiller & Pedersen (2013) are utilized. The factor loadings are depicted in *Figure 1* of the appendix. My weight mechanism is obtained as follows:

I take the factor loadings from Frazzini, Kabiller & Pedersen (2013):

- BaB factor: 0.29
- Value factor: 0.46
- Quality factor: 0.43

I then normalize the factors and obtain their respective weights:

$$w_{BaB} = 0.24576$$

$$w_V = 0.38983$$

$$w_q = 0.36441 \tag{7}$$

Were,

$$W_{BaB} + W_V + W_q = 1 \tag{8}$$

I then compute the final monthly score of individual securities:

$$Ranking_{i,t} = BaB_{Ranking} * w_{BaB} + V_{Ranking} * w_V + q_{Ranking} * w_q \tag{9}$$

Where $Ranking_{i,t}$ is the final ranking of security *i,* at month *t*.

Given the final monthly score, I rank the securities in ascending order, where the lowest scored security is number one in my ranking. I then create a portfolio composed of the top 1% of securities in the dataset, given on the ranking. I invest long in the securities for one month. I calculate the excess monthly returns of the portfolio as follows:

$$R_{P,t} = \frac{1}{N}\sum^{N} r_{E(i,t)} \tag{10}$$

$R_{P,t}$ represents the monthly excess returns of the portfolio. The portfolio returns are calculated as the average sum of excess monthly returns of individual securities over a given month. I rebalance the portfolios every month.

I compare my strategy to two other portfolios, which I view in the next section. First a long-short portfolio, constructed in the same manner as the main strategy. Second, a naïve 1/N portfolio.

4.1. Alternative strategies

The second strategy I implement is a long-short strategy. That is, the same steps as above are reiterated. However, instead of investing only long in the top 1%, I invest short in the bottom 1% as well.

$$R_{P,t}(T) = \frac{1}{N}\sum^{N} r_{E(i,t)}\,(T) \tag{11}$$

$$R_{P,t}(B) = \frac{1}{N}\sum^{N} r_{E(i,t)}(B) \tag{12}$$

Where *T* represent the *top 1%* securities and *B* represents the *bottom 1%* of securities.

The final excess returns of the strategy are computed as follows:

$$P_{\ t} = R_{P,t}(T) - R_{P,t}(B) \tag{13}$$

Where $P_{\ t}$ is the portfolio excess return of the strategy at time *t*.

In addition, a naïve 1/N portfolio is created as follow:

$$R_{naïve,t} = \frac{1}{N}\sum^{N} r_{E(i,t)} \tag{14}$$

Where $R_{naïve,t}$ is the equally weighted sum of excess return of individual securities $r_{E(i,t)}$.

5. Expected Hypothesis

I went through the logic of the strategy and the rationale behind the three factors. Given the novelty of the strategy, I first aim to test it against a naïve 1/N portfolio. I start by looking at the difference in returns between my strategy and the benchmark I picked. I then look at the alpha of my strategy. To that end, I first test the strategy against CAPM. I then continue by looking at its robustness against the FF3 model of Fama & French (1993). I then see the strength of my results against the FF5 (Fama & French, 2015). Knowing that the strategy is the first of its kind, I can see the results going in any direction. However, I expect the strategy to produce an alpha as stocks with characteristics that have proven to produce alpha, in the past, are selected. My results can fail to significantly outperform the naïve portfolio or not. My results can fail to produce a significant alpha or not. I came up with the following hypothesis:

First, I test if:
H0: The top 1% concentrated strategy yields the same results as the 1/N strategy.
HA: The top 1% concentrated strategy yields different results than the 1/N strategy.

Second, I test if;
H0: The long-short strategy yields the same results as the 1/N strategy.
HA: The long-short strategy yields different results than the 1/N strategy.

Third, I test if:
H0: The concentrated strategy yields an alpha equal to zero against the CAPM.
HA: The concentrated strategy yields an alpha different from zero against the CAPM.

Fourth, I test if:
H0: The concentrated strategy yields an alpha equal to zero against the 3-factor model.
HA: The concentrated strategy yields an alpha different from zero against the 3-factor model.

Fifth, I test if:
H0: The concentrated strategy yields an alpha equal to zero against the 5-factor model.
HA: The concentrated strategy yields an alpha different from zero against the 5-factor model.

6. Results

Having described the strategy, its possible outcomes and hypothesis, I can start looking at the results. I begin the section by examining the average monthly excess returns obtained for each of the portfolios created. Furthermore, I explain the presence of extreme returns with market events or outliers. I continue the section by digging into the statistical test. I look at the difference in excess returns and I test for alpha. I finish the section by analyzing the key metrics of my strategy and comments regarding them.

Before starting, I have to note, I controlled for outliers in the sample by excluding securities having a price lower than 5$. The method is frequently used in research to control for outliers (Jagadeesh and Titman 1993). However, I am seeking for undervalued securities in my portfolios. Some omitted securities might have been included in the selection if not discarded by the 5$ control. This limitation can change the direction of my results. Furthermore, when controlling for outliers, I did not control for the liquidity of securities. Illiquid securities with a value of more than 5$ can be included in my analysis. This can further bias my results.

All the returns of the strategy reported are excess returns.

6.1. Concentration strategy

In this section I investigate the results of the factor-based concentration strategy and analyze their key components.

6.1.1. Factor-based concentration strategy – Top 1%

I start to implement the factor-based strategy in February 1995. The sample starts in 1970, 60 months of data are needed before obtaining a beta estimate. Only after computing the first beta value (i.e., in 1975), the first portfolio is constructed out of the ranking. A total of 576 portfolios are constructed until the end of 2022. The average number of securities in the portfolios is 22 holdings.

The average excess returns for the whole sample are:

- 1.36% monthly or
- 16.37% annually.

The monthly excess returns are depicted in the *Figure 2* of the appendix.

Throughout the sample, I observe a maximum return of 0,2644 or 26,44%, in May 1983. No major economic events, able to explain the extreme return, happened during the period. The returns could be the result of outliers. In addition, the minimum return of my strategy reaches -0,1711 (-17.11%) in October 1987. On the 19th of October 1987, “Black Monday” occurred, at which point the stock market crashed unexpectedly. On a single day, the DOW lost 22.6% in value (Trinidad, 2022). The black swan event illustrates that the negative value for my strategy might not be caused by outliers. Other than these two observations, my returns are fairly normal and cause no concern.

6.1.2. Long-Short concentration strategy – Top 1% / Bottom 1%

The long-short concentration strategy is implemented in the same manner as the long only portfolios. The average number of holdings in the portfolios is 44.

The average excess return for the whole sample is:

- 1.33% monthly or
- 16.02% annually.

The monthly returns are depicted in the *Figure 3* of the appendix.

In the figure, I observe three peaks of values. The first one occurs in October 1987, as for the long only strategy. This event is not considered as an outlier in the sample. The negative monthly value for the given month is -0.2743 (-27.43%). In February 2000, I observe a positive 56.96% return (0.5696). This month marks the moment where the FED raised the short-term interest rates. Generally viewed as a negative sign for economic growth, the given news had a negative impact on stock returns (Bebar, 2000). As the strategy captures the downside of the market, it is expected that I profit from such events. However, a 50% increase is rather extreme. That is, the returns observed in the given months might result in part from outliers as well as the short leg of my strategy. The last unexpected peak in the strategy occurs in May 2015. I do not find economic events that could have caused the returns to spike by 45.61% (0.4561) in May 2015. Being nonetheless a year of movement, with different flash crashes, the decline in crude oil prices or the default of Greece on its debt (Mahmudova, 2015). I do not pin down the excessive results to a particular event. That is, outliers might be the reason for the unexpected returns of the given month.

I continue and look at the 1/N naïve strategy.

6.2. 1/N naïve portfolio strategy

When building a trading strategy, there are a few important things to look at. First, does the strategy yields positive and consistent returns through time. Second, does the strategy perform better than an effortless 1/N portfolio. Third, does the strategy yield a significant and robust alpha. I have observed the average excess returns of the strategy. The returns look consistent. I am now interested to see if the strategy performs significantly better than a naïve portfolio. I first look at its construction.

The naïve portfolio is constructed by equally weighting all the securities I have in the sample. Each portfolio has an average of 2200 stocks.

The average excess return for the whole sample is:

- 0.715% per month or
- 8.58% annually.

The monthly returns are depicted in the *Figure 4* of the appendix.

I observe some unexpected returns for the 1/N strategy. The returns correspond to: January 1976, October 1978, October 1987, August 1998, October 2008, April 2009, March 2020 and November 2020. These returns of high magnitude could be considered as outliers. However, when comparing the monthly unexpected returns with the market returns of Kenneth R. French's data library, I see that they are in line with the market. The difference in returns is analyzed below.

6.2.1. Kenneth R. French market returns and the naïve portfolio

I am interested in comparing the returns of the naïve portfolio with the market returns. Having similar excess returns between the two portfolios would indicate that the naïve composition is reliable as a benchmark for the strategies. *Figure 5 of the appendix* exhibits the difference in returns between the naïve portfolio and the excess market returns of Kenneth R. French.

From the observations, I see that the difference between the naïve and market portfolios does not exceed 8%. When looking at the data, the difference in the portfolios is bounded between 0,0717 and -0.0741. A potential explanation for the difference in returns could come from the fact that I use equal weight to construct my portfolio whereas Kenneth R. French uses a value weighted approach to construct his portfolio. Furthermore, some of the divergences in returns can come from my data cleaning process or the method I used to deal with outliers.

From the previous section, I observed eight spikes in the 1/N portfolio. I compare the returns obtained to the market portfolio. I observe a reasonable difference in extreme returns between the market and naïve portfolios. *Table 1 of the appendix* confirms that the extreme results obtained from the strategy are in line with the market returns. Nonetheless, I observe a maximum difference in excess returns of 7.41%. The divergence can be related to the different weighting methods or outliers. To confirm, I test the difference in excess returns of the two portfolios and apply a Newey-West standard errors test. The null hypothesis states that the difference in excess returns between the strategies is zero. In the results, I fail to reject the null that the difference in excess returns is zero (p-value 0.849). The results of the three Newey-West standard errors test I perform are summarized in *Table 1 below.* The full results are depicted in *Table 2 of the appendix.* The test points towards a healthy naïve portfolio construction. Therefore, the naïve strategy is used as a tool of comparison with the other strategies as it provides a reliable benchmark.

Summary table Newey-West standard errors test			
	Coefficient	**Std Error**	**P-value**
1/N vs Market	0,0002	0,001	0,849
Long only vs 1/N	0,0064	0,002	0,002
Long-Short vs 1/N	0,0063	0,002	0,008

Table 1: Summary Newey-West standard errors test

6.3. Strategy comparison

I have analyzed the returns of the different strategies implemented. My next point of interest is to evaluate their relative performances. I look at the difference in returns between the top 1%/long-short compared to the naïve portfolio.

The relative performances of my strategies are evaluated by applying a Newey-West standard errors test with two lags, as in the previous section. The test looks at the average difference in returns between the strategies I implemented. The test controls for autocorrelation and provides p-values for the significance of the difference in excess returns between the strategies.

If I fail to reject the null hypothesis, it indicates that the factor-based concentration strategy fails to outperform the naïve composition. However, if the null is rejected, it indicates a significant difference in returns between the strategies tested.

6.3.1. Long-only Factor-based concentrated strategy and the naïve portfolio

I start to look at the difference in returns between my factor-based strategy and the naïve portfolio. Looking at *Table 1* above[10], a residual average value of 0.0064 is obtained, with a p-value of 0.002. This result indicates the ability of the concentration strategy to significantly outperform the naïve portfolio. That is, the long only strategy yields significantly better returns than a naïve portfolio. With the given results, I can reject the first null hypothesis that the difference in returns between the factor-based strategy and the 1/N strategy is zero. Indeed, from the previous section, I observed that my factor based concentrated portfolio yields an average yearly excess return of 16.37% while the 1/N strategy yields an annual excess return of 8.58%.

6.3.2. Long-short strategy and the naïve portfolio

Table 1 above[11] demonstrates that I can reject the second hypothesis. That is, the average difference in returns between the long-short and the naïve strategy is significantly different from zero. From the coefficient, I observe that the difference is positive (0.0063), pointing towards the outperformance of the long-short strategy compared to the naïve portfolio. The p-value obtained is 0,008. This means that I can reject the null at any conventional level.

6.4. Testing for alpha

I have approached the second step of my analysis and witnessed a significant out-performance from my factor-based strategy. I now go to the third step of the process and examine the alpha extracted. In total, six regressions are covered. Each regression is performed twice, one for the long only and one for the short only strategy. The full results are included in the appendix. *Table 2* below summarizes the results of the three tests performed.

---

[10] A full summary of the results is depicted in *Table 3* of the appendix.

[11] A full summary of the results is depicted in *Table 4* of the appendix.

This section is divided into three parts. I first test for the CAPM model. I continue by using the three-factor model. I finish the section by challenging the alpha with the five-factor model.

Summary table Alpha				
	Coefficient	Std Error	P-value	R-Squared
Long only CAPM	0,011	0,002	0,000	0,126
Long-Short CAPM	0,007	0,003	0,013	0,328
Long only FF3	0,009	0,002	0,000	0,266
Long-Short FF3	0,006	0,002	0,015	0,496
Long only FF5	0,007	0,002	0,001	0,288
Long-Short FF5	0,008	0,003	0,002	0,528

Table 2: Summary table Alpha

6.4.1. CAPM

After running the Newey-West standard errors test, I examine the alpha of my strategy. The alpha of a model represents the excess returns which cannot be explained by the factors I test the model against. In this section, I test my strategy against the excess market returns or CAPM. If a significant alpha is extracted, it shows that the excess returns of the strategy cannot be fully explained by the market. In other words, alpha value represents all the return from the portfolios that the market risk factor is unable to explain.

Looking at the results of the test in *Table 2* above [12], I observe that my strategy has an alpha of 1.06%. This alpha is significant at any conventional level (p-value 0.000). In this case, the market fails to fully explain the returns extracted in my strategy. Looking at the Long-Short strategy versus the market, I observe an alpha of 0.66%[13]. Smaller than for the long only strategy, I can reject the null at a 5% significance level (p-value 0.013). Comparing the two R-squared, the long only strategy has one of 0.126, while the long-short portfolio has an R-squared of 0.328.

The given results show a net performance of my strategy compared to the market returns. I can reject the third null hypothesis.

[12] A full summary of the results is depicted in *Table 5* of the appendix.

[13] A full summary of the results is depicted in *Table 6* of the appendix.

6.4.2. 3-Factor model

Having observed the robustness of the strategy to the market returns, I aim to go a step further. It is known to market participants that anomalies exist. Creating portfolios based on anomalies yields abnormal returns, not explained by the market. Controlling for the factors and testing for alpha confirms the robustness of the strategy to the factors. I start with Fama & French's 3-Factor model (1993). The 3-Factor model considers the size and value factors.

Looking at *Table 2* above[14], I observe an alpha of 0.86% for the long only strategy. The alpha is significant at any conventional level (p-value 0.000). The output shows that the strategy is robust to the 3-factor model and shows that I extract a significant alpha with the portfolio composition. Looking at the long-short strategy[15], the alpha is 0.55% and has a p-value of 0.015. That is, I can reject the null at a 5% significance level. The R-squared of the two strategies are 0.266 for the long only and 0.496 for the long-short portfolio.

I have observed that my results are robust to the 3-Factor model. I can reject the fourth null hypothesis tested. I now look at the 5-Factor model.

6.4.3. 5-Factor model

I conduct a similar test as in the previous section. However, I add a new layer by controlling for two extra market anomalies. In addition to market, size and value, operating profitability and investment are added to the model.

Looking at *Table 2* above[16], I observe that my long only strategy yields an alpha of 0.68% significant at any conventional level (p-value 0.001). Moreover, my strategy produces an alpha after controlling for the five factors. Looking at the long-short strategy, I observe an alpha of 0.79% which is significant at any conventional level (p-value 0.002[17]). The R-squared is 0.288 for the long only and 0.528 for the long-short strategy.

---

[14] A full summary of the results is depicted in *Table 7* of the appendix.

[15] A full summary of the results is depicted in *Table 8* of the appendix.

[16] A full summary of the results is depicted in *Table 9* of the appendix.

[17] A full summary of the results is depicted in *Table 10* of the appendix.

Based on these results, I can reject the last null hypothesis and conclude that my excess returns are robust.

In the previous three tests, I observed that my strategy produces a significant alpha at every conventional level. I shall now tackle the last step of my analysis, which investigates the key metrics of my strategy.

6.5. Key metrics comparison

Having seen that my factor-based strategies produce a significant and robust alpha, I now look at the key metrics of the strategies.

Strategies Key Statistics

Strategy	top_n	long_short	1/n	Market
mean	**0,1637**	**0,1602**	**0,0858**	**0,0821**
std	**0,1770**	**0,2680**	**0,1644**	**0,1562**
min	-0,1711	-0,2743	-0,2635	-0,2324
max	0,2644	0,5696	0,1637	0,1366
median	0,0104	0,0077	0,0098	0,0109
percentile_25	-0,0195	-0,0324	-0,0187	-0,0194
percentile_75	0,0445	0,0498	0,0359	0,0357
Sharpe_ratio	**0,9252**	**0,5977**	**0,5221**	**0,5258**

Table 3: Strategies key statistics

Looking at the comparison of results in *Table 3*, I am especially interested in the Sharpe ratio of my strategy[18]. Indeed, the Sharpe ratio gives us a measure of risk adjusted relative returns. I can see that I obtain a Sharpe ratio of 0.9252. That is, for each unit of risk taken, the strategy can generate an excess return of 0.93. From the four strategies, I observe the long only strategy to yields the best risk adjusted returns. The long-short portfolio exhibits a higher standard deviation for a given excess return, resulting in a weaker Sharpe ratio and thus a lower risk reward strategy. The standard deviation of my strategy lies third (17.70% annually). This result is to be expected, as by constructing a concentrated trading strategy, I expect the risk to be higher than the average market. Looking at the median returns, I observe that the median of my strategy is at 0.0104. This means that 50% of the monthly excess returns lie above or 1.04%.

[18] The results in bold are in annual terms. The results not in bold are denominated on a monthly basis. All the results are in excess returns.

Only the market portfolio of Kenneth R. French has a median average excess return higher than the concentrated strategy.

I have seen and analyzed all my results. I saw that my factor-based strategy significantly outperforms the benchmark portfolio (1/N). In addition, I tested for an alpha against three models. The results are significant and show that my strategy produces an alpha even after being controlled for different factors. Knowing my results and the potential power of my strategy, I have to point out its limitations. Indeed, the strategy outperforms a naïve portfolio and extract an alpha. Caveat and bias could be the driver of my returns and hide a real underperformance. For that reason, in the next section, I cover the limitations of my work.

7. Limitations

My strategy has proven to perform significantly compared to my benchmark and produced an alpha. As positive as my results might be, I must remain cautious and consider the limitations of my work.

In the data section, I covered the different cleaning processes I implemented to obtain clean data that can be deemed ready to work with. I saw that price points were missing (1.08%). I decided to replace the missing values with the previous ones. This transformation had an impact on my dataset and create bias. In addition to that, the quality factor was subject to missing values as well. For the factor, 75,131 values were missing (2.57%). Replacing the given values with the previous ones can have an impact on the data, the ranking of stocks and on the stocks selected for the portfolios. Additionally, replacing the values by the previous ones can impact the data in a significant manner. That is, the quality values are sometimes released once a quarter by companies. If a value is missing and replaced by the previous one, it means that certain securities have the same quality factor for half a year. This might have affected the granularity of the data as well as the precision of the results.

Second, the measure for the factors can impact the strategy excess returns significantly. For example, for the quality factor, I decided to use Novy-Marx's (2013) metric instead of the one of Asness, Frazzini, and Pedersen's (2013). A different measure is used than the one advocated, as no data is available to construct it correctly. However, using a different measure can impact my results.

Third, I have controlled for outliers in the dataset by excluding values under 5$. This transformation allows us to get partly rid of penny stocks. However, by implementing a strategy which aims to invest in cheap undervalued stocks, the method might have erased some opportunities. In addition, when controlling for outliers, volume is often considered. That is, a stock with a low price but high volume and market cap should not be considered as a penny stock. The reason is that high volume usually decreases volatility and thereby the chance of unexpected high fluctuation. In addition, market caps can be high but for a low price; e.g., to make the stock more attractive for investors. The given explanation and way of dealing with outliers might have introduced bias in my dataset. This can undermine my results. Fourth, I observe an R-square of higher magnitude for the long-short strategy. Further analysis should be done to investigate its nature.

Fifth, I do not account for transaction costs. It introduces a bias in my study which overestimates the true returns.

Last, no robustness checks were performed in my analysis. I confirm that my strategy outperforms the benchmark. I see that a significant alpha is produced. Nonetheless, to be certain about the direction of my results, additional tests can be performed. This introduces an extra bias in my analysis.

Having discussed the drawbacks of my research, I can now point towards the room for improvement and what could be done in further research to push my analysis one step further.

8. For Further Research

From the limitation section, different directions could be taken for further research.

In further research, the authors can find a different technique to deal with missing values and compare it to the one I used, to see if different results are obtained. Second, another methodology could be implemented regarding the construction of the factors. Third, outliers could be delt with by considering volume and market cap. Fourth, the ranking of the number could be done in another way. Stocks which are already included in the portfolio of the previous month could obtain a priority ranking. This would decrease the turnover of the strategy. It would as well be a way to reduce costs, as suggested by Barroso & Wang (2021). In addition, it might yield better results as I would invest for a longer term in undervalued securities. Fifth, robustness checks can be performed with the strategy. Testing for the universe of large cap with high liquidity, international evidence or different subsamples can be implemented as robustness tests.

Sixth, the strategy can be implemented to other markets, either to compare it to the analysis I have already performed, or to do a cross comparison between markets other than the US. Seventh, the strategy seems to be profitable, in theory. However, I have no proof in practice. The strategy can be implemented in the real market. A paper strategy can first be tested to then implement it with capital, if proven to be successful. Eighth, additional factors can be found and added to the analysis. In that case, the base strategy can be compared with the augmented one. Ninth, transaction costs can be included. Last, monthly data was used. Granularity might have been lost by doing so. The next step can be to implement the strategy with daily observations.

9. Conclusion

To conclude, through my research, I aim to create a concentration trading strategy offering a reliable and consistent way of investing for retail investors.

I started my analysis by looking at traditional portfolio theory. Portfolio theory provides a guide on the way rational investors should behave to maximize their returns given a level of risk. From that observation, I discovered that practice diverges from theory. I learned from different studies that investors seem to deviate from the optimal choice. Indeed, evidence point towards the preference of retail to opt for concentrated portfolios. This is what let me to wonder if such a concentrated portfolio could be implemented and produce an appealing risk-return trade-off. To my knowledge, studies have investigated the matter but do not offer straightforward solutions for retail investors, leaving a gap in the literature. Based on my observations, I came up with the following research question which I aim to answer through my analysis. Investment

Performance of a Concentrated Factor-Based Strategy: to what extent is it possible to produce an alpha with the creation of a factor-based trading strategy?

My main research question led me to ask myself the following sub questions:

- Is it possible to replicate Warren Buffett's strategy in a simple way?
- Can a factor-based concentrated strategy beat a naïve portfolio?
- Can I extract significant returns with an undiversified concentrated portfolio?
- Can I extract an alpha with a factor-based concentration strategy?

To find answers to my questions, I proceeded as follows.

Having a clear idea of the gap between theory and practice, I started my analysis by wondering about a tool to create a strategy. In my quest, I looked at Frazzini, Kabiller, and Pedersen's analysis (2013). Their research seeks to explain the Warren Buffett's performance with factors. They link the performance of the manager with a few metrics and the use of the leverage. In their analysis, the authors do not apply their findings to create a strategy.

This pushed me to take the process one step further and use their results to create a factor-based trading strategy. To that end, I picked three factors: quality, value and beta, significant in Frazzini, Kabiller, and Pedersen (2013), to build indicators to select stocks. The three factors are known to perform well in predicting stocks returns, which confirmed my choices.

To create the strategy, I took the three factors selected and came up with a ranking. My ranking aggregates the individual scores of securities based on the three factors. Given the triple ranking, one for each factor, a final score is obtained. The final score offers an indication on which securities to include in the portfolios.

I create one factor-based long only concentrated strategy, one factor-based long-short concentration strategy and one naïve portfolio to be used as a benchmark. I find supportive evidence that the strategies yield statistically different returns compared to a naïve portfolio. I obtain an annual excess return of 16.37% for the long only and 8.58% for the naïve portfolio. The difference in returns is significant at any conventional level (p-value 0.002). The same is observed for the long-short strategy (p-value 0.008), where I extract an annual excess return of

16.02%. In addition to that, I test my strategies for alpha. First, the returns are challenged by the market. I extract significant and positive alpha for the long and long-short strategies. Second, the two strategies are put to the test against the 3-factor model. In the second test, both strategies have a positive and significant alpha. Third, the 5-factor model is put in use to gauge the robustness of my returns against two extra anomalies. For the 5-factor model, I obtain an alpha of 0.0068, for the long strategy, significant at any conventional level (p-value 0.001). The long-short strategy produces an alpha of 0.0079, significant at any conventional level as well (p-value 0.002). For my results, I can conclude that an alpha is extracted by applying the strategies I introduced. However, I have to keep in mind the different limitations I faced during the project. The results I obtained should be taken with caution and a new analysis should be conducted to confirm my findings.

In conclusion, I have desined a factor-based trading strategy, which aims to create a concentrated portfolio, easily replicable by retail investors. I have extracted an alpha, robust to different factor models. The results obtained significantly outperform the benchmark used. In addition, a Sharpe ratio of 0.9252 is found for the short strategy.

10. Appendix

Table 1: Market vs 1/N Excess Returns.

Difference between naïve and market excess returns. The months with the highest difference in returns are depicted. The dates are analyzed in text.

Date	Market	1/N	Difference
Jan-76	0,1216	0,1607	0,0391
Oct-78	-0,1191	-0,1657	-0,0466
Oct-87	-0,2324	-0,2635	-0,0311
Aug-98	-0,1608	-0,1654	-0,0046
Oct-08	-0,1723	-0,1856	-0,0133
Apr-09	0,1018	0,1637	0,0619
Mar-20	-0,1339	-0,2080	-0,0741
Nov-20	0,1247	0,1603	0,0356

Table 1: Market vs 1/N excess returns

Table 2: Newey-West standard errors test of the difference between the naïve portfolio and the market excess returns.

The results of the Newey-West standard errors test show that I fail to reject the null hypothesis. The null states that the difference in excess returns between the naïve portfolio and the market are zero. The test provides proof that both portfolios behave in a similar fashion. The test confirms that the naïve portfolio behaves according to the market. The test confirms that I can used the 1/N strategy as a reliable benchmark.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     diff_1_over_N_and_market   R-squared:                      -0.000
Model:                                  OLS   Adj. R-squared:                 -0.000
Method:                       Least Squares   F-statistic:                       nan
Date:                      Sun, 26 Mar 2023   Prob (F-statistic):                nan
Time:                              11:15:52   Log-Likelihood:                 1441.2
No. Observations:                       575   AIC:                            -2880.
Df Residuals:                           574   BIC:                            -2876.
Df Model:                                 0
Covariance Type:                        HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const_1        0.0002      0.001      0.191      0.849      -0.002       0.002
==============================================================================
Omnibus:                       23.061   Durbin-Watson:                   1.843
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.494
Skew:                           0.247   Prob(JB):                     3.59e-10
Kurtosis:                       4.254   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 3: Newey-West standard errors test of the difference between the long only portfolio and the naïve portfolio excess returns.

The results of the Newey-West standard errors test show that I reject the null hypothesis. The null states that the difference in excess returns between the long only and naïve portfolio are zero. The test provides proof that the long portfolio has a difference in excess return with the benchmark which is statistically different from zero. This difference is significant at any conventional level.

```
                               OLS Regression Results
==============================================================================
Dep. Variable:     diff_Top_1_and_1_over_N   R-squared:                       0.000
Model:                                 OLS   Adj. R-squared:                  0.000
Method:                      Least Squares   F-statistic:                       nan
Date:                     Sun, 26 Mar 2023   Prob (F-statistic):                nan
Time:                             11:15:51   Log-Likelihood:                 911.40
No. Observations:                      575   AIC:                            -1821.
Df Residuals:                          574   BIC:                            -1816.
Df Model:                                0
Covariance Type:                       HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const_1        0.0064      0.002      3.116      0.002       0.002       0.010
==============================================================================
Omnibus:                       31.588   Durbin-Watson:                   2.064
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               69.226
Skew:                           0.299   Prob(JB):                     9.28e-16
Kurtosis:                       4.591   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 4: Newey-West standard errors test of the difference between the long-short portfolio and the naïve portfolio excess returns.

The results of the Newey-West standard errors test show that I reject the null hypothesis. The null states that the difference in excess returns between the long-short and naïve portfolio are zero. The test provides proof that the long-short portfolio has a difference in excess return with the benchmark which is statistically different from zero. This difference is significant at any conventional level.

```
                                  OLS Regression Results
====================================================================================
Dep. Variable:     diff_Long_Short_and_1_over_N   R-squared:                       0.000
Model:                                      OLS   Adj. R-squared:                  0.000
Method:                           Least Squares   F-statistic:                       nan
Date:                          Sun, 26 Mar 2023   Prob (F-statistic):                nan
Time:                                  11:15:52   Log-Likelihood:                 820.98
No. Observations:                           575   AIC:                            -1640.
Df Residuals:                               574   BIC:                            -1636.
Df Model:                                     0
Covariance Type:                            HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const_1        0.0063      0.002      2.634      0.008       0.002       0.011
==============================================================================
Omnibus:                      379.703   Durbin-Watson:                   2.008
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6885.033
Skew:                           2.614   Prob(JB):                         0.00
Kurtosis:                      19.126   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 5: Long only portfolio and the CAPM

The table reports the risk adjusted returns – alpha – of the long only portfolio when tested against the market, the CAPM model. I obtain a significant positive value for alpha of 1.06%. The alpha is significant at any conventional level.

```
                               OLS Regression Results
==================================================================================
Dep. Variable:     Monthly_Reurns_Top_1_per   R-squared:                       0.126
Model:                                  OLS   Adj. R-squared:                  0.125
Method:                       Least Squares   F-statistic:                     56.57
Date:                      Wed, 22 Mar 2023   Prob (F-statistic):           2.11e-13
Time:                              10:00:55   Log-Likelihood:                 932.95
No. Observations:                       575   AIC:                            -1862.
Df Residuals:                           573   BIC:                            -1853.
Df Model:                                 1
Covariance Type:                        HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0106      0.002      4.673      0.000       0.006       0.015
mkt_rf         0.4026      0.054      7.521      0.000       0.298       0.508
==============================================================================
Omnibus:                       67.994   Durbin-Watson:                   1.746
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              128.514
Skew:                           0.710   Prob(JB):                     1.24e-28
Kurtosis:                       4.829   Cond. No.                         22.2
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 6: Long-short portfolio and the CAPM

The table reports the risk adjusted returns – alpha – of the long-short portfolio when tested against the market, the CAPM model. I obtain a significant positive value for alpha of 0.66%. The alpha is significant at 5% level.

```
                               OLS Regression Results
====================================================================================
Dep. Variable:     Monthly_Returns_Long_Short   R-squared:                       0.328
Model:                                    OLS   Adj. R-squared:                  0.326
Method:                         Least Squares   F-statistic:                     294.2
Date:                        Wed, 22 Mar 2023   Prob (F-statistic):           1.56e-53
Time:                                10:00:56   Log-Likelihood:                 770.62
No. Observations:                         575   AIC:                            -1537.
Df Residuals:                             573   BIC:                            -1529.
Df Model:                                   1
Covariance Type:                          HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0066      0.003      2.475      0.013       0.001       0.012
mkt_rf         0.9810      0.057     17.153      0.000       0.869       1.093
==============================================================================
Omnibus:                      343.103   Durbin-Watson:                   1.962
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4915.020
Skew:                           2.348   Prob(JB):                         0.00
Kurtosis:                      16.531   Cond. No.                         22.2
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 7: Long only portfolio and the 3-Factor model

The table reports the risk adjusted returns – alpha – of the long only portfolio when tested against the 3-Factor model. I obtain a significant positive value for alpha of 0.86%. The alpha is significant at any conventional level.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Monthly_Reurns_Top_1_per   R-squared:                       0.266
Model:                                  OLS   Adj. R-squared:                  0.262
Method:                       Least Squares   F-statistic:                     48.33
Date:                      Wed, 22 Mar 2023   Prob (F-statistic):           7.56e-28
Time:                              10:00:56   Log-Likelihood:                 983.14
No. Observations:                       575   AIC:                            -1958.
Df Residuals:                           571   BIC:                            -1941.
Df Model:                                 3
Covariance Type:                        HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0086      0.002      4.500      0.000       0.005       0.012
mkt_rf         0.3771      0.047      7.954      0.000       0.284       0.470
SMB            0.5476      0.082      6.697      0.000       0.387       0.708
HLM            0.4543      0.071      6.395      0.000       0.315       0.594
==============================================================================
Omnibus:                       61.104   Durbin-Watson:                   1.919
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              116.394
Skew:                           0.645   Prob(JB):                     5.31e-26
Kurtosis:                       4.787   Cond. No.                         37.3
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 8: Long-short portfolio and the 3-Factor model

The table reports the risk adjusted returns – alpha – of the long-short portfolio when tested against the market, the 3-Factor model. I obtain a significant positive value for alpha of 0.55%. The alpha is significant at 5% level.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Monthly_Return_Long_Short   R-squared:                       0.496
Model:                                   OLS   Adj. R-squared:                  0.493
Method:                        Least Squares   F-statistic:                     150.0
Date:                       Wed, 22 Mar 2023   Prob (F-statistic):           1.08e-71
Time:                               10:00:56   Log-Likelihood:                 853.57
No. Observations:                        575   AIC:                            -1699.
Df Residuals:                            571   BIC:                            -1682.
Df Model:                                  3
Covariance Type:                         HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0055      0.002      2.427      0.015       0.001       0.010
mkt_rf         0.8093      0.059     13.735      0.000       0.694       0.925
SMB            1.1131      0.154      7.245      0.000       0.812       1.414
HLM            0.0564      0.092      0.615      0.538      -0.123       0.236
==============================================================================
Omnibus:                      285.442   Durbin-Watson:                   2.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2488.600
Skew:                           2.004   Prob(JB):                         0.00
Kurtosis:                      12.370   Cond. No.                         37.3
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 9: Long only portfolio and the 5-Factor model

The table reports the risk adjusted returns – alpha – of the long only portfolio when tested against the 5-Factor model. I obtain a significant positive value for alpha of 0.68%. The alpha is significant at any conventional level.

```
                                 OLS Regression Results
===================================================================================
Dep. Variable:     Monthly_Reurns_Top_1_per   R-squared:                       0.288
Model:                                  OLS   Adj. R-squared:                  0.282
Method:                       Least Squares   F-statistic:                     31.23
Date:                      Wed, 22 Mar 2023   Prob (F-statistic):           4.00e-28
Time:                              10:00:57   Log-Likelihood:               991.85
No. Observations:                       575   AIC:                          -1972.
Df Residuals:                           569   BIC:                          -1946.
Df Model:                                 5
Covariance Type:                        HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0068      0.002      3.446      0.001       0.003       0.011
Mkt_100        0.4120      0.048      8.526      0.000       0.317       0.507
SMB_100        0.6423      0.073      8.791      0.000       0.499       0.786
HML_100        0.2140      0.083      2.565      0.010       0.050       0.378
RMW_100        0.2909      0.085      3.418      0.001       0.124       0.458
CMA_100        0.2841      0.122      2.326      0.020       0.045       0.523
==============================================================================
Omnibus:                       72.499   Durbin-Watson:                   1.940
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              159.825
Skew:                           0.698   Prob(JB):                     1.97e-35
Kurtosis:                       5.173   Cond. No.                         79.2
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Table 10: Long-short portfolio and the 5-Factor model

The table reports the risk adjusted returns – alpha – of the long-short portfolio when tested against the market, the 5-Factor model. I obtain a significant positive value for alpha of 0.79%. The alpha is significant at any conventional level.

```
                                 OLS Regression Results
=====================================================================================
Dep. Variable:     Monthly_Returns_Long_Short   R-squared:                       0.528
Model:                                    OLS   Adj. R-squared:                  0.524
Method:                         Least Squares   F-statistic:                     110.7
Date:                        Wed, 22 Mar 2023   Prob (F-statistic):           1.37e-81
Time:                                10:00:57   Log-Likelihood:               872.63
No. Observations:                         575   AIC:                          -1733.
Df Residuals:                             569   BIC:                          -1707.
Df Model:                                   5
Covariance Type:                          HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0079      0.003      3.121      0.002       0.003       0.013
Mkt_100        0.7869      0.052     15.098      0.000       0.685       0.889
SMB_100        0.9084      0.094      9.701      0.000       0.725       1.092
HML_100       -0.1521      0.122     -1.244      0.214      -0.392       0.088
RMW_100       -0.6580      0.156     -4.205      0.000      -0.965      -0.351
CMA_100        0.2724      0.175      1.560      0.119      -0.070       0.615
==============================================================================
Omnibus:                      261.703   Durbin-Watson:                   1.972
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2060.892
Skew:                           1.831   Prob(JB):                         0.00
Kurtosis:                      11.521   Cond. No.                         79.2
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Figure 1: Buffett's Alpha (Frazzini, Kabiller & Pedersen, 2013) Regression Output.

In the figure I observe the output of the regression made by Frazzini, Kabiller & Pedersen (2013). The factors picked to create the portfolios are highlighted. The factors loading are selected and normalize to create portfolio weight.

	Berkshire stock 1976 - 2011			13F portfolio 1980 - 2011			Private Holdings 1984 - 20011		
Alpha	**12.1%**	**9.2%**	6.3%	**5.3%**	3.5%	0.3%	5.6%	4.6%	4.9%
	(3.19)	(2.42)	(1.58)	(2.53)	(1.65)	(0.12)	(1.35)	(1.08)	(1.09)
MKT	**0.84**	**0.83**	**0.95**	**0.86**	**0.86**	**0.98**	**0.40**	**0.40**	**0.39**
	(11.65)	(11.70)	(10.98)	(21.55)	(21.91)	(20.99)	(5.01)	(5.01)	(3.94)
SMB	**-0.32**	**-0.32**	-0.15	**-0.18**	**-0.18**	0.00	**-0.29**	**-0.29**	**-0.31**
	-(3.05)	-(3.13)	-(1.15)	-(3.14)	-(3.22)	(0.02)	-(2.59)	-(2.53)	-(2.17)
HML	**0.63**	**0.38**	**0.46**	**0.39**	**0.24**	**0.31**	**0.39**	0.28	0.27
	(5.35)	(2.79)	(3.28)	(6.12)	(3.26)	(4.24)	(3.07)	(1.89)	(1.81)
UMD	0.06	-0.03	-0.05	-0.02	**-0.08**	**-0.10**	0.09	0.04	0.05
	(0.90)	-(0.40)	-(0.71)	-(0.55)	-(1.98)	-(2.66)	(1.13)	(0.52)	(0.55)
BAB		**0.37**	**0.29**		**0.22**	**0.15**		0.16	0.17
		(3.61)	(2.67)		(4.05)	(2.58)		(1.40)	(1.41)
QMJ			**0.43**			**0.44**			-0.05
			(2.34)			(4.55)			-(0.24)
R2 bar	0.25	0.27	0.28	0.57	0.58	0.60	0.08	0.08	0.08

Figure 2: Monthly Excess Returns Top 1%.

Monthly excess returns of the long only strategy. Depicted are the excess returns of the strategy. I observe outlier for different point in time. They are described and analyze in text.

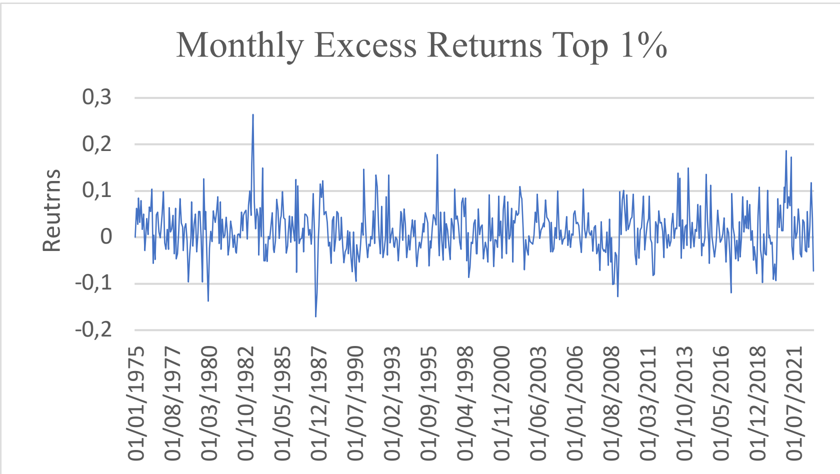

Figure 2: Monthly excess returns Top 1%

Figure 3: Monthly Excess Returns Long-Short.

Monthly excess returns of the long-short strategy. Depicted are the excess returns of the strategy. I observe outlier for different point in time. They are described and analyze in text.

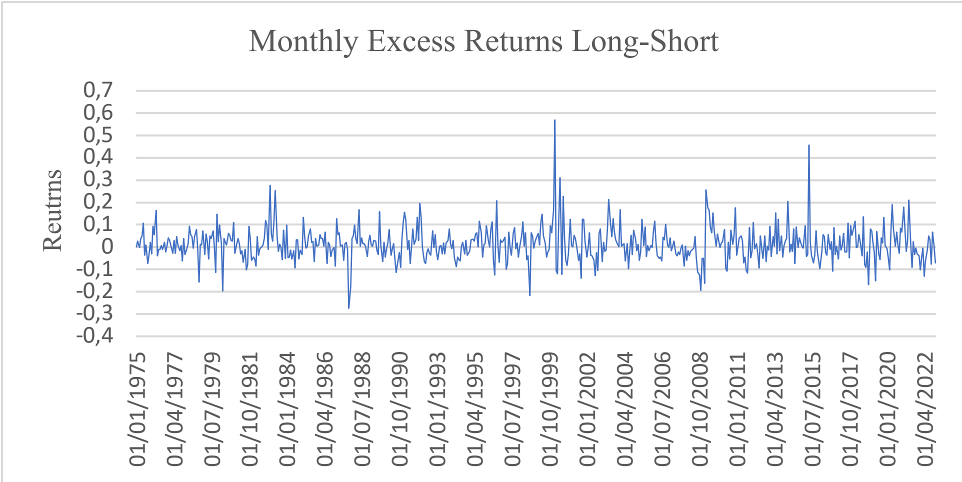

Figure 3: Monthly excess returns Long-Sort

Figure 4: Monthly Excess Returns 1/N.

Monthly excess returns of the naïve portfolio. Depicted are the excess returns of the strategy. I observe outlier for different point in time. They are described and analyze in text.

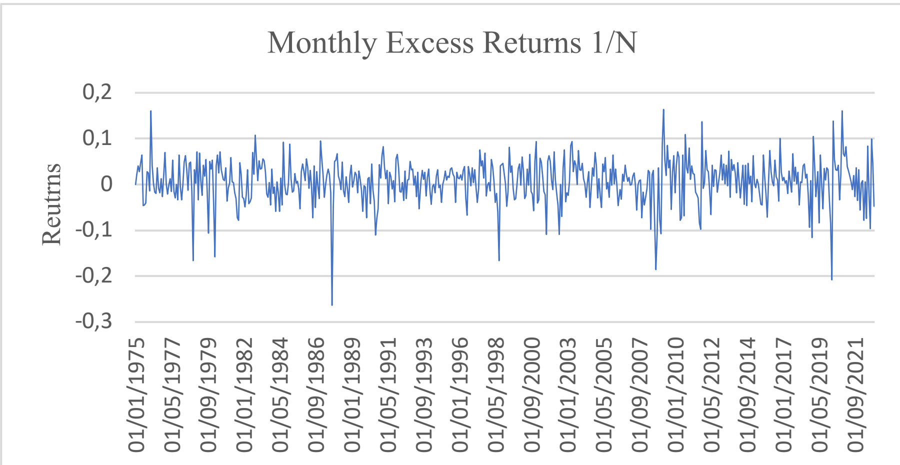

Figure 4: Monthly excess returns 1/N

Figure 5: 1/N Residual Returns to Market.

Difference in excess returns between the naïve portfolio and the market returns are plotted. The 12 months moving average is plotted. The figure is constructed to observe if the naïve portfolio behaves in similar fashion as the market. Seeing relatively similar behavior of returns point toward a reliable construction of the naïve portfolio. The figure aims to confirm that the naïve portfolio can be used as a reliable benchmark.

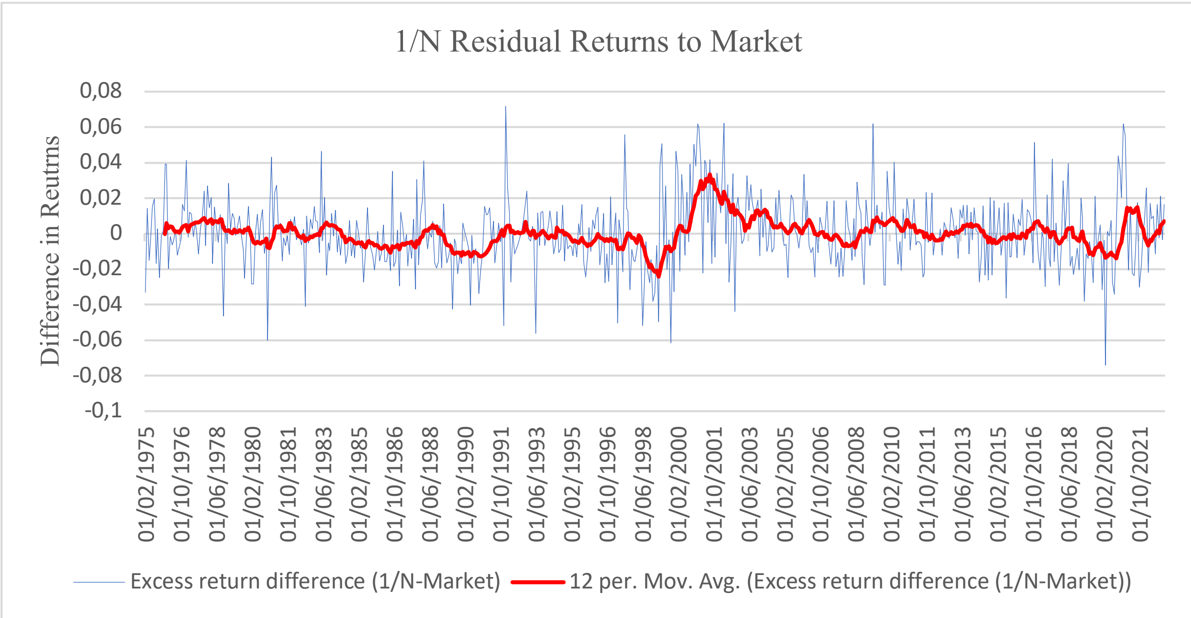

Figure 5: 1/N Residuals to market returns

Figure 6: Cumulative Returns

Plot of the cumulative returns of the three strategies. I observe a net outperformance of the long only/long-short strategies over the naïve portfolio. The total cumulative returns reach 769.10% for the long only portfolio, 765.41% for the long-short portfolio and 403.56% for the naïve portfolio.

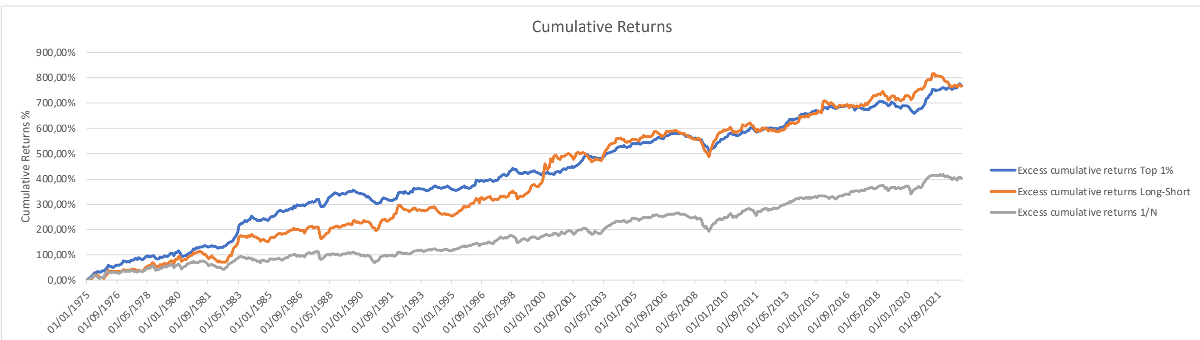

Figure 6: Cumulative Returns

11.References

Agarwal, V., Fos, V., & Jiang, W. (2013). Inferring Reporting-Related Biases in Hedge Fund Databases from Hedge Fund Equity Holdings. *Management Science*, *59*(6), 1271–1289. https://doi.org/10.1287/mnsc.1120.1647

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. *Cambridge University Press EBooks*, 294–305. https://doi.org/10.1017/cbo9780511809477.022

Anderson, A. (2008). Trading and Under-Diversification. In *Social Science Research Network*. Social Science Electronic Publishing. https://doi.org/10.2139/ssrn.871435

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2015). Quality Minus Junk. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2312432

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and Momentum Everywhere. *Journal of Finance*, *68*(3), 929–985. https://doi.org/10.1111/jofi.12021

Barber, B. M., & Odean, T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *Journal of Finance*, *55*(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Barber, B. M., & Odean, T. (2001). Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics*, *116*(1), 261–292. https://doi.org/10.1162/003355301556400

Barber, B. M., & Odean, T. (2013). The Behavior of Individual Investors. *Handbook of the Economics of Finance*, 1533–1570. https://doi.org/10.1016/b978-0-44-459406-8.00022-6

Barberis, N., & Thaler, R. H. (2002). *A Survey of Behavioral Finance*. https://doi.org/10.3386/w9222

Barroso, P. (2017). Lest we forget: Using Out-Of-Sample Errors in Portfolio Optimization. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2771664

Barroso, P., & Wang, H. (2021). What Explains Price Momentum and 52-Week High Momentum When They Really Work? *Social Science Research Network*. https://doi.org/10.2139/ssrn.3716786

Bebar, J. (2000, February 2). *U.S. stocks end mixed - Feb. 2, 2000*. CNN. Retrieved March 15, 2023, from https://money.cnn.com/2000/02/02/markets/markets_newyork/

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. In *Econometrica* (Vol. 22, Issue 1, p. 23). Wiley-Blackwell. https://doi.org/10.2307/1909829

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, *45*(3), 444. https://doi.org/10.1086/295472

Bodie, Z., Kane, A., & Marcus, A. (2012). *Essentials of Investments* (9th ed.). McGraw-Hill Education.

Bodie, Z., Kane, A., & Marcus, A. J. (2018). *Essentials of Investments* (11th ed.). McGraw-Hill Education.

Boyle, P. P., Garlappi, L., Uppal, R., & Wang, T. (2009). Keynes Meets Markowitz: The Tradeoff Between Familiarity and Diversification. *Social Science Research Network*. https://doi.org/10.2139/ssrn.1365014

Campbell, J. (2006). Household Finance. *Journal of Finance*, *61*(4), 1553–1604. https://doi.org/10.1111/j.1540-6261.2006.00883.x

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Center for Research in Security Prices. (2023). *CRSP/Compustat Merged Database | CRSP - The Center for Research in Security Prices*. Retrieved February 14, 2023, from https://www.crsp.org/products/research-products/crspcompustat-merged-database

Center for Research in Security Prices, LLC. (2020). Data Descriptions Guide CRSP US Stock & US Index Databases. In *CRSP*. Retrieved January 7, 2023, from https://www.crsp.org/files/data_descriptions_guide_0.pdf

Chen, G., Kim, K. A., Nofsinger, J. R., & Rui, O. M. (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. In *Journal of Behavioral Decision Making* (Vol. 20, Issue 4, pp. 425–451). Wiley-Blackwell. https://doi.org/10.1002/bdm.561

Choi, N. Y., Fedenia, M., Skiba, H., & Sokolyk, T. (2017). Portfolio concentration and performance of institutional investors worldwide. *Journal of Financial Economics*, *123*(1), 189–208. https://doi.org/10.1016/j.jfineco.2016.09.007

Columbia University. (2014). *Warren Edward Buffet*. Retrieved February 15, 2023, from https://c250.columbia.edu/c250_celebrates/remarkable_columbians/warren_edward_buffett.html

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *Journal of Finance*, *52*(3), 1035–1058. https://doi.org/10.1111/j.1540-6261.1997.tb02724.x

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/*N*Portfolio Strategy? *Review of Financial Studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Dorn, D., & Huberman, G. (2005). Talk and Action: What Individual Investors Say and What They Do. *European Finance Review*, *9*(4), 437–481. https://doi.org/10.1007/s10679-005-4997-z

Døskeland, T., & Hvide, H. K. (2011). Do Individual Investors Have Asymmetric Information Based on Work Experience? In *Journal of Finance* (Vol. 66, Issue 3, pp. 1011–1041). Wiley-Blackwell. https://doi.org/10.1111/j.1540-6261.2011.01658.x

Driessen, J., & Laeven, L. (2007). International portfolio diversification benefits: Cross-country evidence from a local perspective. *Journal of Banking and Finance*, *31*(6), 1693–1712. https://doi.org/10.1016/j.jbankfin.2006.11.006

Ekholm, A., & Maury, B. (2013). Stock Picking and Firm Performance. *Social Science Research Network*. https://doi.org/10.2139/ssrn.1572604

Evans, J. S. O., & Archer, S. L. (1968). DIVERSIFICATION AND THE REDUCTION OF DISPERSION: AN EMPIRICAL ANALYSIS*. *Journal of Finance*, *23*(5), 761–767. https://doi.org/10.1111/j.1540-6261.1968.tb00315.x

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *38*(1), 34. https://doi.org/10.1086/294743

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405x(93)90023-5

Fama, E. F., & French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, *51*(1), 55–84. https://doi.org/10.1111/j.1540-6261.1996.tb05202.x

Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, *18*(3), 25–46. https://doi.org/10.1257/0895330042162430

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252. https://doi.org/10.1016/j.jfineco.2018.02.012

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Frazzini, A., Kabiller, D. G., & Pedersen, L. H. (2013). Buffett's Alpha. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3197185

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005

Fulkerson, J. A., & Riley, T. B. (2019). Portfolio concentration and mutual fund performance. *Journal of Empirical Finance*, *51*, 1–16. https://doi.org/10.1016/j.jempfin.2019.01.006

Goetzmann, W. N., & Kumar, A. (2004). Equity Portfolio Diversification. *Social Science Research Network*. https://doi.org/10.2139/ssrn.627321

Graham, B., Dodd, D., & Dodd, D. L. F. (1934). *Security Analysis: The Classic 1934 Edition*. McGraw Hill Professional.

Griffin, J. H., & Xu, J. (2009). How Smart Are the Smart Guys? A Unique View from Hedge Fund Stock Holdings. *Review of Financial Studies*, *22*(7), 2531–2570. https://doi.org/10.1093/rfs/hhp026

Grinblatt, M., & Keloharju, M. (2009). Sensation Seeking, Overconfidence, and Trading Activity. In *Journal of Finance* (Vol. 64, Issue 2, pp. 549–578). Wiley-Blackwell. https://doi.org/10.1111/j.1540-6261.2009.01443.x

Gruber, M. J. (1996). Another Puzzle: The Growth in Actively Managed Mutual Funds. *Journal of Finance*, *51*(3), 783–810. https://doi.org/10.1111/j.1540-6261.1996.tb02707.x

Hatchondo, J. C. (2008). ASYMMETRIC INFORMATION AND THE LACK OF PORTFOLIO DIVERSIFICATION*. *International Economic Review*, *49*(4), 1297–1330. https://doi.org/10.1111/j.1468-2354.2008.00513.x

Hou, K., Xue, C., & Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *Review of Financial Studies*, *28*(3), 650–705. https://doi.org/10.1093/rfs/hhu068

Huberman, G. (2001). Familiarity Breeds Investment. *Review of Financial Studies*, *14*(3), 659–680. https://doi.org/10.1093/rfs/14.3.659

Huij, J., & Derwall, J. (2009). Global Equity Fund Performance, Portfolio Concentration, and the Fundamental Law of Active Management. *Social Science Research Network*. https://doi.org/10.2139/ssrn.1456352

Hvidkjaer, S. (2008). Small Trades and the Cross-Section of Stock Returns. *Review of Financial Studies*, *21*(3), 1123–1151. https://doi.org/10.1093/rfs/hhn049

Ivković, Z., Sialm, C., & Weisbenner, S. J. (2008). Portfolio Concentration and the Performance of Individual Investors. *Journal of Financial and Quantitative Analysis*, *43*(3), 613–655. https://doi.org/10.1017/s0022109000004233

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance*, *48*(1), 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jensen, M. C. (1968). THE PERFORMANCE OF MUTUAL FUNDS IN THE PERIOD 1945-1964. *Journal of Finance*, *23*(2), 389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x

Kacperczyk, M., Sialm, C., & Zheng, L. S. (2005). On the Industry Concentration of Actively Managed Equity Mutual Funds. *Journal of Finance*, *60*(4), 1983–2011. https://doi.org/10.1111/j.1540-6261.2005.00785.x

Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of US households. *Journal of Economic Behavior and Organization*, *27*(1), 87–96. https://doi.org/10.1016/0167-2681(95)00006-e

Lakonishok, J. (2004). Value and Growth Investing: Review and Update. *Financial Analysts Journal*, *60*(1), 71–86. https://doi.org/10.2469/faj.v60.n1.2593

Levy, H., & Sarnat, M. (1970). International Diversification of Investment Portfolios. *American Economic Review*, *60*(4), 668–675.

Lhabitant, F. (2017). *Portfolio Diversification*. Elsevier.

Lintner, J. (1965a). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, *47*(1), 13. https://doi.org/10.2307/1924119

Lintner, J. (1965b). Security Prices, Risk, and Maximal Gains From Diversification. *Journal of Finance*, *20*(4), 587. https://doi.org/10.2307/2977249

Liu, H. (2008). A New Explanation for Underdiversification. *Journal of Economic Literature*. https://www.researchgate.net/profile/Anjan-Thakor/publication/229053921_A_New_Explanation_for_Underdiversification/links/0c96052980060641b4000000/A-New-Explanation-for-Underdiversification.pdf

Liu, H. (2014). Solvency Constraint, Underdiversification, and Idiosyncratic Risks. *Journal of Financial and Quantitative Analysis*, *49*(2), 409–430. https://doi.org/10.1017/s0022109014000271

Liu, J., Stambaugh, R. F., & Yuan, Y. (2018). Absolving beta of volatility's effects. *Journal of Financial Economics*, *128*(1), 1–15. https://doi.org/10.1016/j.jfineco.2018.01.003

Mahmudova, A. (2015, December 22). *The events that rocked financial markets in 2015*. MarketWatch. Retrieved March 16, 2023, from https://www.marketwatch.com/story/the-events-that-rocked-financial-markets-in-2015-2015-12-22

Malkiel, B. G. (1995). Returns from Investing in Equity Mutual Funds 1971 to 1991. *Journal of Finance*, *50*(2), 549–572. https://doi.org/10.1111/j.1540-6261.1995.tb04795.x

Markowitz, H. M. (1952). Portfolio Selection. *Journal of Finance*, *7*(1), 77. https://doi.org/10.2307/2975974

Marston, R. C. (2011). *Portfolio Design: A Modern Approach to Asset Allocation*. John Wiley & Sons.

Martin, G. S., & Puthenpurackal, J. (2007). Imitation is the Sincerest Form of Flattery: Warren Buffett and Berkshire Hathaway. *Social Science Research Network*. https://doi.org/10.2139/ssrn.806246

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, *41*(5), 867. https://doi.org/10.2307/1913811

Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics*, *8*(4), 323–361. https://doi.org/10.1016/0304-405x(80)90007-0

Moore, D. A., & Healy, P. M. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295x.115.2.502

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, *34*(4), 768. https://doi.org/10.2307/1910098

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28. https://doi.org/10.1016/j.jfineco.2013.01.003

Odean, T. (1999). Do Investors Trade Too Much? *The American Economic Review*, *89*(5), 1279–1298. https://doi.org/10.1257/aer.89.5.1279

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. In *Journal of Financial Economics* (Vol. 80, Issue 3, pp. 563–602). Elsevier BV. https://doi.org/10.1016/j.jfineco.2005.05.003

Polkovnichenko, V. (2004). Household Portfolio Diversification: A Case for Rank Dependent Preferences. *Social Science Research Network*. https://doi.org/10.2139/ssrn.293831

Rosenberg, B., Reid, K. B., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16. https://doi.org/10.3905/jpm.1985.409007

Sapp, T., & Yan, X. (2008). Security Concentration and Active Fund Management: Do Focused Funds Offer Superior Performance? *The Financial Review*, *43*(1), 27–49. https://doi.org/10.1111/j.1540-6288.2007.00185.x

Scheinkman, J. A., & Xiong, W. (2003). Overconfidence and Speculative Bubbles. *Social Science Research Network*. https://doi.org/10.2139/ssrn.298865

Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. *Management Science*, *9*(2), 277–293. https://doi.org/10.1287/mnsc.9.2.277

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, *19*(3), 425. https://doi.org/10.2307/2977928

Statman, M., & Scheid, J. (2008). Correlation, Return Gaps, and the Benefits of Diversification. *The Journal of Portfolio Management*, *34*(3), 132–139. https://doi.org/10.3905/jpm.2008.706250

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148. https://doi.org/10.1016/0001-6918(81)90005-6

Treynor, J., & Black, F. (1973). How to Use Security Analysis to Improve Portfolio Selection. *The Journal of Business*, *46*(1). https://www.jstor.org/stable/2351280

Treynor, J. L. (1962). Jack Treynor's "Toward a Theory of Market Value of Risky Assets." *Social Science Research Network*. https://doi.org/10.2139/ssrn.628187

Trinidad, C. (2022, December 20). *Black Monday*. Corporate Finance Institute. Retrieved March 10, 2023, from https://corporatefinanceinstitute.com/resources/equities/black-monday/

Uppal, R., & Wang, T. (2003). Model Misspecification and Underdiversification. *Journal of Finance*, *58*(6), 2465–2486. https://doi.org/10.1046/j.1540-6261.2003.00612.x

Van Horne, J. C., Blume, M. E., & Friend, I. (1975). THE ASSET STRUCTURE OF INDIVIDUAL PORTFOLIOS AND SOME IMPLICATIONS FOR UTILITY FUNCTIONS. *Journal of Finance*, *30*(2), 585–603. https://doi.org/10.1111/j.1540-6261.1975.tb01833.x

Veldkamp, L., & Van Nieuwerburgh, S. (2010). Information Acquisition and Under-Diversification. *The Review of Economic Studies*, *77*(2), 779–805. https://doi.org/10.1111/j.1467-937x.2009.00583.x

Wharton Research Data Services. (n.d.-a). *Overview of CRSP U.S. Stock Database*. © 1993-2023 Wharton Research Data Services. Retrieved February 14, 2023, from https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/crsp/stocks-and-indices/overview-crsp-us-stock-database/

Wharton Research Data Services. (n.d.-b). *Overview of CRSP/Compustat Merged Data*. © 1993-2023 Wharton Research Data Services. Retrieved February 15, 2023, from https://wrds-www.wharton.upenn.edu/pages/support/data-overview/ccm-overview-crspcompustat-merged-database/

Wharton Research Data Services. (2020). CRSP/Compustat Merged Database Guide. In *Wharton Research Data Services*. Retrieved February 15, 2023, from https://wrds-www.wharton.upenn.edu/documents/402/CRSP-Compustat_Merged_Database_Data_Guide_9efDcmD.pdf

Winters, M. (2023, March 1). *Warren Buffett's "secret sauce" for investing success: Be "business pickers" not "stock pickers."* CNBC. Retrieved March 12, 2023, from https://www.cnbc.com/2023/03/01/warren-buffetts-secret-sauce-for-investing-success.html